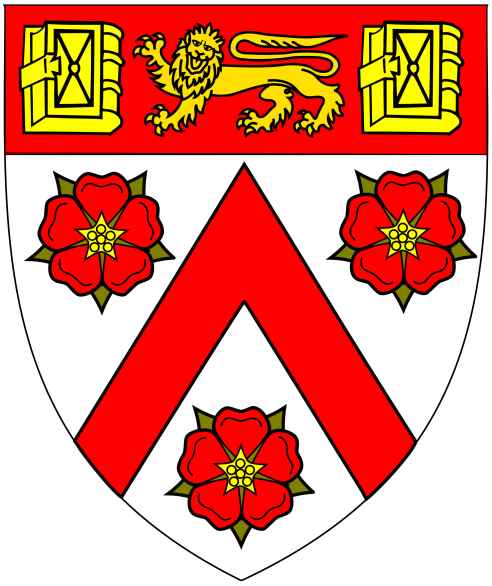
Population dynamics and population genetics of the Critically Endangered Raso lark: implications for conservation

Elisa Gwenda Godelieve Dierickx
Trinity College, University of Cambridge



This dissertation is submitted for the degree of Doctor of Philosophy
August 2017

Table of contents

Preface

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface or in the Acknowledgements and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the Biology Degree Committee.

Acknowledgements

The past four years as a PhD student have been the experience of a lifetime. I am greatly indebted to my supervisor, Dr. Michael Brooke, for giving me the unique opportunity to work on the Raso lark and in Cape Verde, a country whose nature conservation projects I have particularly at heart. Thank you Mike for being so generous with your time and for your patience - especially when trying to catch the skittish Raso larks of 2014!

I would like to thank my two advisors, Prof. Nick Davies and Dr. Nick Mundy, for their help and words of encouragement over the course of my PhD. The Zoology Department was a wonderful place to work. I am particularly grateful to Prof. Chris Jiggins and the members of the Butterfly Genetics Group for their invaluable help. Sarah Barker guided me through the preparation of my RAD libraries. Without Simon Martin and Ana Pinharanda Chapter 6 would have been a daunting task: I have learned very much from them. Thank you very much to Camille Roux for helping me contextualize the results of Chapter 6. The Behavioural Ecology and Conservation Science Groups have provided very helpful comments about my work over the course of my PhD. “Thesis Writing Club” with Gabriel Jamie, Ben Jarrett and Jack Thorley was a great source of feedback, motivation, brainstorming and good laughs. I am also much indebted to Rob Robinson for his invaluable help with the MARK analyses.

In the “other Cambridge,” at Harvard, Prof. Scott Edwards taught my first ornithology course when I was an undergraduate and has continued to advise me and host me in his research group since then. Thank you Scott, I greatly value your mentorship. I am also grateful to the members of the Edwards Lab for helpful discussions over the years, most of all to Allison Shultz, who taught me more than anyone else in science, was extremely generous with her time, and on whom I could always count for moral support. Thank you very much also to Simon Yung Wa Sin for his work on the skylark genome and for commenting on the two genetics chapters in this thesis. Finally, I would like to acknowledge the support of Harvard’s Research Computing - I don’t know what I would have done without my dear friend Odybot, their live chat system!

Many thanks to Prof. Yang Liu, Dr. Nick Horrocks, Dr. Michael Brooke, the Harvard MCZ and Pieter van Veelen for sharing their painstakingly-collected lark samples with me, and to Edwin Winkel for letting me use his beautiful photographs of

Raso larks. I am also very grateful for the generous financial support that I have received from the Peter Scott Studentship for Conservation Biology, Vocatio Belgium, Trinity College and the Zoology Department. Thank you to Trinity College for providing the most beautiful working and living conditions one could dream of.

My friends are wonderful too: Carmen and Kate, whom I missed very much in the office after they graduated, Arthur, Ting, Marius, Chantal, Melissa and Natasha for being so supportive while I was on Raso, and all of them, which includes Cherry and Elise, for being my family away from home. The RPG group was great at making me forget about work and stress for a few hours every week.

Finally, I would like to thank my family, whom I love more than anything: my father, who is the best coach I could wish for, my mother, because, as she (correctly) points out, “mothers are always right,” my sister Marion, who finished her own PhD earlier this year and whom I admire very much, and my brother Jan, whose wise advice I should follow more often. My love also goes to my grandfather, a keen birder and bird breeder, who inspired me to become an ornithologist the day he gave me my first chickens, Pim, Bleekje, Prijs and Bruintje.

Summary

The Raso lark is a Critically Endangered bird endemic to the islet of Raso, Cape Verde. This thesis investigates two phenomena that particularly put the species at risk: its extreme fluctuations in population size, and its potentially very low genetic diversity arising from small population size and severe past population contraction. More specifically, two chapters estimate year-to-year survival and explore the factors - environmental and individual - that influence it, while two other chapters examine the lark's genetic characteristics compared to its two continental closest relatives, including phylogenetic relationships and levels of genetic diversity. The conclusion of the thesis then uses these results to make recommendations for the conservation of the Raso lark. Each of the data chapters is summarized below:

Chapter 3 estimates adult survival in the Raso lark and tests whether it could be linked to two population phenomena observed in the field: a highly variable population size and a male-biased sex ratio in certain years. Using a dataset spanning 10 years, I estimated survival for both sexes to fluctuate between 0.76 and 0.94 over this period. This is much higher than the survival rate of its closest relative, the skylark. I also found strong evidence for survival fluctuating over time and differing between males and females (with males having higher survival until 2011, at which point the trend inverted), which could play a role in the aforementioned population size fluctuations and male-biased sex ratio, respectively.

Chapter 4 aims at understanding which factors shape survival in the Raso lark. Two types of variables were considered: year-dependent (rainfall, population size, population mean clutch size) and individual-dependent (age, body size characters, size ratio with mate, Ase18 genotype). Amongst the year-dependent variables, only same-year rainfall impacted survival, with a 13% decrease in survival in the wettest year compared to the driest year, making it the most likely explanation for the inter-annual fluctuations in survival found in Chapter 3. Results also hint at some of the individual factors - morphological measurements and Ase18 genotype - influencing survival. The picture that emerges is that of a species whose life history strategy is to invest heavily in maintenance and survival, but less into fecundity, which stands in sharp contrast with the mainland-dwelling skylark. This is consistent with the theory that island birds generally have slower life history strategies than their continental counterparts.

Chapter 5 determines the precise relationship between members of the *Alauda* clade, resolving a node on the phylogenetic tree of all larks that the study by Alström et al. (2013) was unable to resolve. My RADseq results indicate that the Raso lark and the skylark are sister species, and that the Oriental lark is likely to be a subpopulation, or maybe a subspecies, of the skylark.

Chapter 6 compares the population genetics of the Raso lark with those of the skylark. In particular, it estimates the genetic diversity of the Raso lark and investigates the drivers behind it. I found unexpectedly high nucleotide diversity in the Raso lark, and explain this by showing that the population contraction that the species underwent was recent enough for most of the diversity to still be present. Moreover, 16% of the Raso lark genome has levels of heterozygosity on average 6.6 times higher than elsewhere on the genome, likely due to suppressed recombination and the existence of a neo-sex chromosome in larks. Despite this, I found high levels of relatedness and of linkage disequilibrium in the Raso lark, two clear genetic signs that it underwent a severe population contraction several centuries ago.

Chapter 1: Introduction

Boyd Alexander (1873-1910) was the English ornithologist, explorer and army officer who, in 1897, discovered the Raso lark, during one of his voyages around Africa.

Abstract

Island species have suffered 89% of all avian extinctions, despite only representing 20% of all bird species. Underlying this phenomenon is the vulnerability of island species to alien invasive species, which were involved in 53% of island extinctions since 1500, but "only" in 37.7% of continental extinctions. Other "Achilles' heels" of island species include threats that are linked to the intrinsic geographical characteristics of islands such as isolation, proximity to ocean and small area. In the future, climate change may disproportionally impact island species. The isolation and often small size of island populations reduce gene flow, which can cause different genetic hazards, including inbreeding depression. Stochastic environmental events, habitat destruction or resource depletion can also potentially harm island species much more than continental species, since the former may be unable to disperse away from the events. The Raso lark, a bird endemic to the tiny islet of Raso in Cape Verde, is vulnerable to many of these threats and, as a result, is listed as Critically Endangered. This thesis investigates two phenomena that threaten the species: its extreme fluctuations in population size, and its potentially very low genetic diversity arising from small population size and severe past population contraction. More specifically, after an overview of the general methods in chapter 2, chapters 3 and 4 estimate year-to-year survival and explore the factors - environmental and individual - that influence it, while chapters 5 and 6 examine the lark's genetic characteristics compared to its two continental closest relatives, including phylogenetic relationships and levels of genetic diversity. A final chapter then uses these results to make recommendations for the conservation of the Raso lark.

Insular versus continental avian extinctions

It is well documented that island birds have suffered far more species extinctions than continental birds (Figure 1.1). At least 150 bird species have gone extinct since 1500, the date of the earliest extinction documented on the IUCN Red List. This corresponds to an average extinction rate of 0.30 species per year, increasing to 0.56 species per year in the years since 1900 (Butchart et al. 2006). Three species are known or suspected to have gone extinct in the wild since 2000: Spix's macaw *Cyanopsitta spixii*, the Hawaiian crow *Corvus hawaiiensis* and the Po'ouli *Melamprosops phaeosoma*. Of these extinct species, ≈ 89% were insular, despite the fact that worldwide more than 80% of bird species are continental (Johnson and Stattersfield 1990; Manne, Brooks and Pimm 1999; Butchart et al. 2006; BirdLife International 2011). In sum, the extinction rate of island bird species in modern times is around 40 times higher than that of continental bird species (Johnson and Stattersfield 1990).

As illustrated by Figure 1.2, extinctions particularly took place in Hawaii (27 species), Mauritius (18 species), New Zealand (14 species), Réunion (11 species) and St. Helena (9 species) (Butchart et al. 2006). The Atlantic Ocean was home to four of the 97 extinctions since 1600 (Figure 1.2). Currently 50% of the Atlantic's endemic avian species are threatened, a higher proportion than other parts of the world, where it ranges from 10% in New Guinea and Melanesia to 38% in the Pacific (Johnson and Stattersfield 1990). Passerines represent 49 of the 151 avian extinctions. Of these, three were continental, and 46 were insular (Butchart et al. 2006).

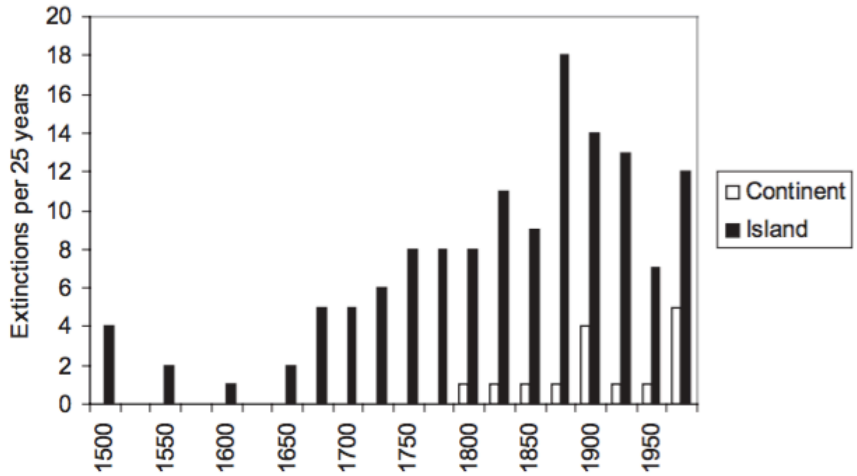


Figure 1.1 Number of probable bird species extinctions on continents and on islands (copied from Butchart et al. 2006).

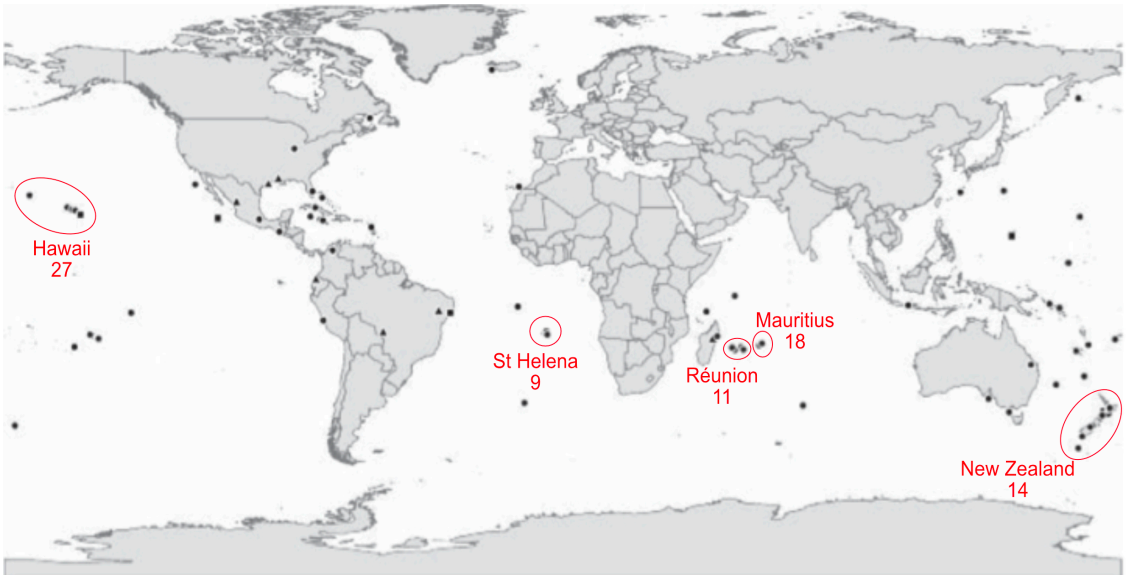


Figure 1.2 Distribution of probable avian extinctions since 1500. The dots indicate the location of the last known record of each species. Circled in red are the areas of the world that suffered the most extinctions. The numbers correspond to the number of species gone extinct in each of these areas (adapted from Butchart et al. 2006).

What are the reasons behind higher extinction rates amongst island birds than amongst continental birds? The most critical one is the introduction of invasive aliens, thought to have been involved in 77 - or 51% - of all avian extinctions since 1500. Invasive aliens disproportionately affect island species, causing 53% of island extinctions but "only" 37.5% of continental extinctions (Butchart et al. 2006). Mammals, and more specifically cats *Felix catus*, rodents of the order Rodentia, dogs *Canis familiaris* and pigs *Sus domesticus*, threaten the most species overall (Johnson and Stattersfield 1990; Doherty et al. 2016). Island species are thought to be more vulnerable to invasive alien species and pathogens for two reasons, both resulting from the fact that they evolved in isolation: island ecosystems' relatively few plants, herbivores, carnivores and decomposers performing vital processes (Lowe et al. 2000) renders them more vulnerable to break-down because of the lack of functional redundancy, and island species often present island tameness and ecological naïveté (Hagman, Phillips and Shine 2009; Lee, Wood and Rogers 2010; Carthey and Banks 2014; Cooper, Pyron and Garland 2014), including lack of resistance and immunity to diseases (Lowe et al. 2000; Wyatt et al. 2008).

The threat posed by invasive aliens to endemic species can take different shapes: predation, competition, habitat destruction and/or disease. Invasive alien predators were involved in at least 56 extinctions (Butchart et al. 2006), and potentially up to 87

(Doherty et al. 2016). One example is the disappearance of the Stephen's Island wren *Traversia lyalli* when cats were introduced onto the island in 1894 (Butchart et al. 2006). Diseases caused by pathogens were responsible for 20 extinctions, 16 of which were in Hawaii. Many Hawaiian birds died when avian malaria *Plasmodium relictum* and, at an earlier stage, its vector the southern house mosquito *Culex quinquefasciatus* were introduced onto the islands (Butchart et al. 2006; Samuel et al. 2011). Habitat destruction, often by grazers such as goats *Capra hircus*, sheep *Ovis aries* and rabbits of the family Leporidae, caused 10 extinctions (Butchart et al. 2006). As a more specific example, the introduction of crazy ants *Anoplolepsis gracilipes* on Christmas Island threatens the forest canopy which serves as the sole nesting site of the Endangered Abbott's booby *Sula abbotti* worldwide (Lowe et al. 2000). Competition was at the base of six extinctions (Butchart et al. 2006) and threatens more species today. For example, introduced rodents in New Zealand compete for food with the Endangered North Island brown kiwi *Apteryx mantelli* (Shapiro 2005). Finally, an invasive alien species can have effects that cascade through whole ecosystems, with disastrous consequences. For example, over the past 200 years, after their introduction to Australia, feral cats and red foxes *Vulpes vulpes* have caused the extinction of two thirds of Australian digging mammal species through predation. The absence of these endemic digging mammals greatly reduced the disturbance in the topsoil, which led to the creation of impoverished landscapes with very little organic matter accumulation and very low rates of seed germination (Doherty et al. 2016).

Climate change can also disproportionately affect insular species (Parmesan 2006; Bellard et al. 2012; Ferreira et al. 2016; Slavenko et al. 2016). Climate change can negatively impact both continental and insular species in numerous and complex ways, ranging from increased atmospheric CO_2 concentrations to increased ocean pH to seasonal temperature changes (Parmesan 2006; Bellard et al. 2012). Most of these phenomena disproportionately affect island species because of the inherent geographical characteristics of islands. While a continental species can "respond to climate change by shifting its climatic niche along three non-exclusive axes - time (e.g. phenology), self (e.g. physiology) and space (e.g. range)" (Bellard et al. 2012), island species only have very limited access to the latter. Because of the smaller areas and isolation of islands, insular species are often unable to shift their ranges in response to climate change-related phenomena such as rising atmospheric temperatures or decreased rainfall (Parmesan 2006; Bellard et al. 2012). Furthermore, because islands have a disproportionately large number of species living or nesting in close proximity to the

ocean, extreme weather events such as storms or hurricanes are particularly dangerous to them (Dierickx et al. 2015).

These intrinsic geographical characteristics of islands (isolation, proximity to ocean and small area) underlie the remaining threats to insular species. In some cases these threats are island-specific, in others the characteristics of islands amplify them. Isolation may make gene exchange difficult, which, coupled to the often small size of island populations, can increase the risk of genetic depletion and inbreeding (Frankham 1998; Ferreira et al. 2016). Even flying animals like birds and bats can exhibit high philopatry and limited dispersal abilities (e.g. Chaverri and Kunz 2011). An example that has received publicity recently is that of the last woolly mammoth *Mammuthus primigenius* population - 300 individuals surviving on Wrangel Island - that is thought to have ultimately disappeared because of an accumulation of bad mutations that led to "genomic meltdown" (Rogers and Slatkin 2017). Other known threats to biodiversity such as stochastic environmental events, habitat destruction or resource depletion can potentially have much stronger effects on island species than on continental species due to the inability of the former to migrate (Bellard et al. 2012). And finally, it is noteworthy that 48.7% of extinctions were the result of multiple threats acting together (Johnson and Stattersfield 1990; Butchart et al. 2006).

Could these geographical characteristics of islands also *favour* insular species in any way? In recent years, while the continental extinction rate appears to be increasing, that of island extinctions seems to be slowing (Butchart et al. 2006; BirdLife International 2011). Indeed, when surveying currently endangered bird species - as opposed to extinct species - one finds that only 39% (402 species) are restricted to islands (Johnson and Stattersfield 1990), despite representing 59% (659 species) of all bird species (Manne, Brooks and Pimm 1999). This phenomenon is probably due to increased habitat destruction on the mainland, to the past purging of the most vulnerable island species, to the fact that many of the potential introductions of invasive aliens have already happened and, in a more positive vein, to successful island conservation measures (Butchart et al. 2006; McCreless et al. 2016; BirdLife International 2011).

Today islands can - after a pest eradication campaign - serve as alien-free refuges for endangered species. Their smaller size makes such eradication programs more manageable, and their isolation renders re-colonization by alien species less likely. Added to the fact that islands are self-contained ecosystems, their small size also makes habitat restoration more achievable. Their isolation renders the mitigation of threats such as hunting or tourism more manageable than on the mainland through, for

example, the creation of insular reserves that are difficult to access. Isolation may also offer islands protection against epidemics and pathogens. For these reasons, New Zealand's Department of Conservation puts considerable effort in the management of 220 of its offshore islands, including habitat restoration, species reintroductions and the removal of alien invasive species. Over 100 of these islands are now pest-free thanks to the Department's efforts (Department of Conservation 2016).

Cape Verde - its geography

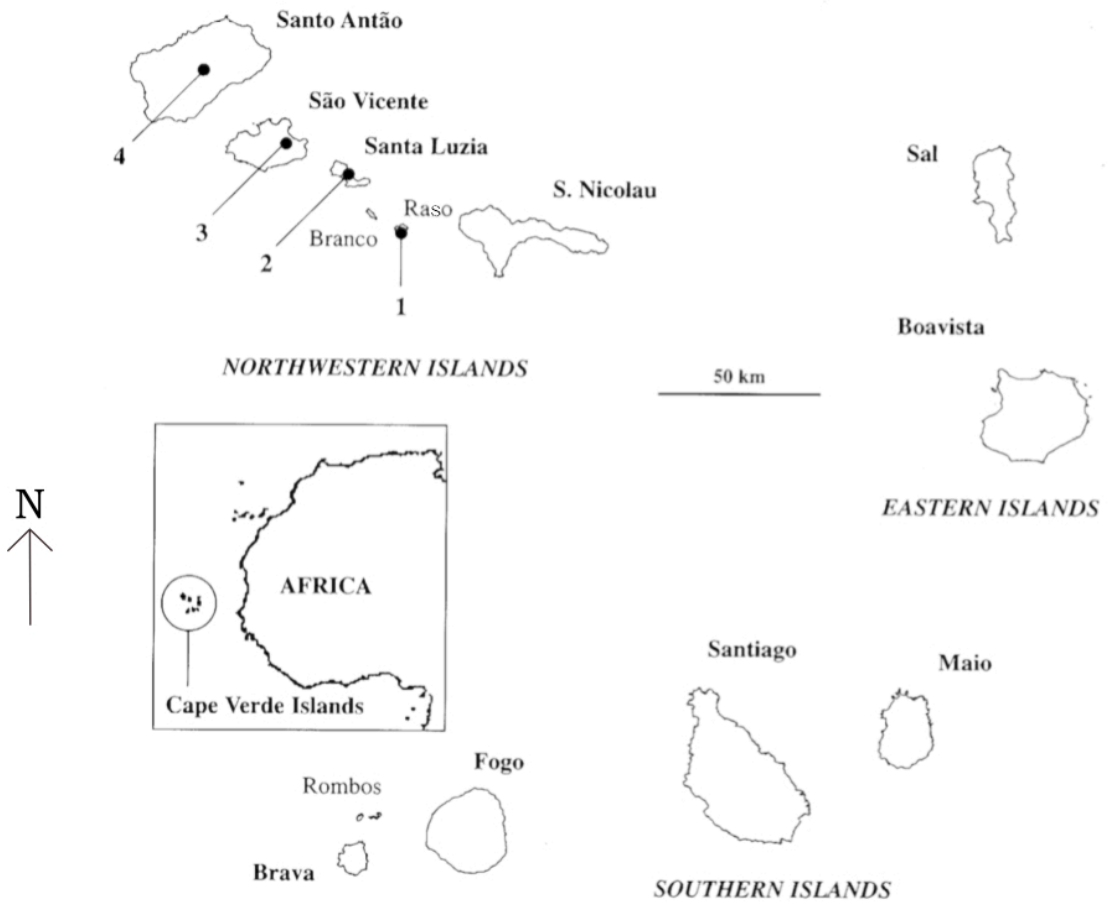


Figure 1.3 Map of the Cape Verde Islands. Numbers 1-4 indicate the Raso lark's range until human settlement in the archipelago during the 15th century, according to the fossil record. Number 1 corresponds to Raso Island, the only place in the world where the species is currently found (adapted from Mateo et al. 2009).

Cape Verde is an archipelago consisting of nine inhabited islands and a number of smaller, uninhabited islands. It lies in the Atlantic Ocean, about 500km west of Senegal (Figure 1.3). The islands are between 8 million years old in the west and 20 million years old in the east. They were formed through volcanic activity above a deep

mantle plume (Pim et al. 2008); today only the Fogo Island volcano remains active. Apart from a few rainy days between September and December, Cape Verde is a very dry and arid country - a feature that made a great impression on Charles Darwin when he first laid eyes on it during his voyage on the Beagle, on 16 January, 1832:

> The neighbourhood of Porto Praya [on Santiago Island], viewed from the sea, wears a desolate aspect. The volcanic fire of past ages, and the scorching heat of a tropical sun, have in most places rendered the soil sterile and unfit for vegetation […] The island would generally be considered as very uninteresting; but to any one accustomed only to an English landscape, the novel prospect of an utterly sterile land possesses a grandeur which more vegetation might spoil. A single green leaf can scarcely be discovered over wide tracts of the lava plains […] It rains very seldom, but during a short portion of the year heavy torrents fall, and immediately afterwards a light vegetation springs out of every crevice. This soon withers; and upon such naturally-formed hay the animals live. At the present time it has not rained for an entire year. The broad, flat-bottomed, valleys, many of which serve during a few days only in the season as a watercourse, are clothed with thickets of leafless bushes. Few living creatures inhabit these valleys.

Cape Verde is regularly hit by terrible droughts that can last up to 18 years. In the 1940s, tens of thousands of people died in a famine caused by the lack of rainfall. The 1970s also saw a ten-year drought (Donald and Brooke 2006). It is unclear whether the current lack of vegetation and aridity were caused by human settlement of the archipelago (Requedaz 2014). The archipelago was uninhabited until the Portuguese discovered it in 1456 and colonized it from 1462 onwards (Donald and Brooke 2006).

It is on one of the archipelago's smaller islets, Raso (Figures 1.3 and 1.4), that one can find the Raso lark *Alauda razae* (Figure 1.5). Raso is about 7km^2 (Mateo et al. 2009), uninhabited, and currently free of mammals. Just like the rest of Cape Verde, Raso is extremely dry and arid - there is no water to be found on the island - and Darwin's description very much applies to it. The island is surrounded by cliffs, and has two main landscape types. The first is found on the lower parts of the island, and consists of flat, rocky plains that average 25m in altitude and are traversed by *ribeiras* (dry river beds that only rarely fill with water). These plains are covered in lava rocks and scarce, often dry vegetation. The second type of landscape consists of rocky hills

culminating at 164m (Ratcliffe, Monteiro and Hazevoet 1999), valleys and plateaux. These parts of the island have even less vegetation than the plains (Figure 1.4).

Figure 1.4 Landscapes on Raso Islet, from left to right and from top to bottom: the plains in a rainier year (2013), the same plains in a drier year (2014), a view onto the plains from the plateau, and the plateau with Branco and Santa Luzia Islands in the distance. © Elisa Dierickx

The Raso lark, a bird endemic to Cape Verde

Figure 1.5 Raso lark female. © Edwin Winkel

In 1897, a young British explorer named Boyd Alexander, pictured on the title page of this chapter, landed on Raso Islet. There he discovered a new lark species, which he named *Spizocorys razae*, placing it in a genus that today contains seven African lark species - but no longer the Raso lark (Alexander 1898; Donald and Brooke 2006). Over the years, the species discovered by Alexander was successively re-classified in the genera *Razocorys* (of which it was the only species), C*alandrella*, *Alaudala* and, finally, *Alauda* (Donald and Brooke 2006; Alström et al. 2013). Its closest relatives are the skylark *Alauda arvensis* and the Oriental lark *Alauda gulgula,* but between the three species the node of the phylogenetic tree is not yet resolved (Alström et al. 2013; Figure 5.1).

There are at least two possible explanations for the evolutionary path separating the Raso lark from its two closest relatives, probably at some point during the last seven million years (Alström et al. 2013). The first one is that *Alauda razae* on Raso is the last remaining population of a species that was previously far more common across Africa and disappeared from the continental mainland because of climatic changes. The second one is that the Raso lark evolved *in situ* in Cape Verde. When, possibly during one of the Pleistocene Ice Ages, the Palearctic fauna was found further south in what is now the Sahara, European skylarks or their ancestors could have colonized Cape Verde.

Then, when the ice retreated and the Sahara became desert again, they would have retreated northwards and left some individuals behind in Cape Verde, individuals that evolved into the species found there today (Donald and Brooke 2006).

Until the 1980s, very little was known about the Raso lark's ecology and behaviour. Ornithologists only occasionally and irregularly visited the island until Michael Brooke started ringing individuals every year starting in 2002 (Ratcliffe, Monteiro and Hazevoet 1999; Donald and Brooke 2006; Table 1.1). From that year onwards, population size estimates became more precise and reliable: before then, it is possible that surveyors might have missed the non-breeding flocks on the plateaux, biasing the population size estimates downwards. Nevertheless, it seems that the Raso lark population has fluctuated greatly, with as few as 20 pairs remaining at the beginning of the 1980s (Ratcliffe, Monteiro and Hazevoet 1999; Donald and Brooke 2006), before increasing over the next decades and reaching an estimated peak of 1558 individuals in 2011 (Brooke and Finnie 2012). The most recent population size estimate, from December 2016, was 818 individuals (Brooke and Gregory 2016; Table 1.1).

As population size estimates became more precise, more and more became known about the ecology of the species as well. The Raso lark is a sexually dimorphic species (the males are larger than the females with minimal overlap of wing and bill measurements) that breeds irregularly, following the similarly irregular rainfalls (Donald and Brooke 2006). When breeding is possible, most individuals live and forage in pairs on the plains of Raso (Figures 1.4 and 2.1), with paired males spending a significant amount of time mate guarding, and unpaired males defending their territories (Donald and Brooke 2006). When not breeding, the birds are mainly found foraging in large, skittish flocks on the plateaux of the island (Figure 1.4). Raso larks feed on the bulbs of nutsedges *Cyperus bulbosus* and *Cyperus cadamosti*; males, in particular, even dig deep holes to reach them (Figure 1.6). They also consume invertebrates that they find on vegetation, on the ground surface or under stones. A faecal analysis showed that the larks eat vegetable matter, Lepidoptera and Coleoptera larvae, snails thought to be marine gastropods, and other invertebrates (Donald and Brooke 2006). It is thought that Raso larks get most of their water from their food, although in drought conditions birds are often seen approaching the shoreline at sunset, maybe to drink seawater from the tidal pools.

Table 1.1 Scientific visits to Raso and Raso lark population size estimates 1965-2016 (Ratcliffe, Monteiro and Hazevoet 1999; Brooke and Bolton 2002; Brooke and Hille 2003; Brooke and Gregory 2016).

Year	Population size
1965	< 50 pairs
1968	< 40 pairs
1977	20 pairs
1981	20 pairs
1981 (2nd visit)	20 pairs
1985	> 150 pairs
1986*	200 pairs
1988	75-100 birds
1988 (2nd visit)	250 birds
1989	200 birds
1990	250 birds
1992	250 birds
1998	92 birds
2001	128-138 birds
2002	80-100 birds
2003	80 birds
2004	57 birds
2005	132 birds
2006	140 birds
2007	159 birds
2008	184 birds
2009	193 birds
2010	486 birds
2011	1558 birds
2012	1546 birds
2013	1314 birds
2014	1170 birds
2015	900 birds
2016	818 birds

* First systematic count

Figure 1.6 A Raso lark digging ground in 2014, a relatively dry year. © Elisa Dierickx

The Raso lark has few predators: only two pairs of neglected kestrels *Falco neglectus* and one pair of brown-necked ravens *Corvus ruficollis* live on the island. Additionally, the Cape Verde giant gecko *Tarentola gigas*, also Endangered and endemic to Raso and Branco, predates its nests. Only the gecko is thought to have any significant impact on the population dynamics of the Raso lark (Donald et al. 2003; Donald and Brooke 2006).

If alien predators were introduced to Raso, the impact on the lark would likely be severe - as it has already been for many island bird species. Indeed, fossils of Raso larks were found on the islands of Santa Luzia (35km^2), São Vicente (227km^2) and Santo Antão (779km^2), proving that the species was much more widespread in the past (Figure 1.3) (Mateo et al. 2009). During the Pleistocene, Santa Luzia, São Vicente, Raso and Branco (3km^2) were all connected, forming, together with now-submerged zones between the islands, a super-island of several thousand square kilometres. The island of Santo Antão was also very close to this super-island, separated by a channel of about 12km and presumably within the reach of dispersing larks (Donald and Brooke 2006; Mateo et al. 2009). Importantly, the Raso lark fossils on these islands were found in deeper levels, levels that predate human arrival - and their cats, dogs and rodents - in

Cape Verde. Their abrupt disappearance from the fossil record and consequent absence in the higher levels shows that the Raso lark's extinction from Santa Luzia, São Vicente and Santo Antão was very fast once humans colonized them (Mateo et al. 2009). Today, Raso is the last of the larger islets in Cape Verde that is still cat and rat-free, and also the last one where *Alauda razae* remains. In this thesis, this event will be referred to as the lark's "population contraction" or "population reduction," while the shorter-term fluctuations shown in Table 1.1 will be called "bottlenecks."

Another threat to the Raso lark, as well as to many island species in the future, is climate change. As an arid country in the Sahel region and a small island state, Cape Verde is doubly at risk (Figure 1.7). The average annual temperature has increased by 0.6°C since 1960, a trend that is projected to continue this century. Extreme weather events have become more frequent. Rising sea levels and coastal hazards are particularly dangerous to low-lying islands. Because of its volcanic soils, steep terrain, limited vegetation cover and irregular rainfall, water runoff to the sea during the rainy season is severe, and fresh water is scant. This situation could further deteriorate, as future rainfall patterns are uncertain: some models predicted an increase in rainfall, while others suggested the opposite (Ministry of Environment, Housing and Territory Planning of Cape Verde 2011). Since the Raso lark needs rainfall for breeding, reduced rainfall could be very harmful to the species.

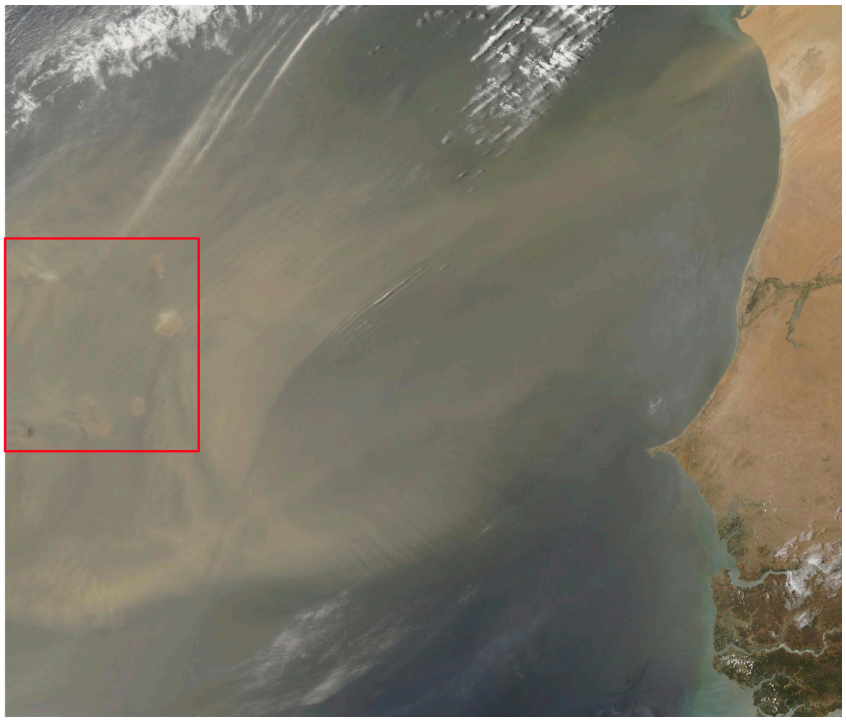


Figure 1.7 A sand storm carries dust from the Sahara over to Cape Verde (red square), illustrating Cape Verde's climatic link to the Sahel region (CIA 2016).

Potentially inimical to the species' persistence are the genetic bottlenecks through which it has passed. The maintenance of genetic diversity is crucial for fitness and survival, both at the individual level and at the species level (Brook et al. 2002; O'Grady et al. 2006; Traill, Bradshaw and Brook 2007). A species with a small effective population size is subject to three types of genetic risk. The first is inbreeding depression, which is reduced fitness due to the increase in homozygotes caused by mating between relatives (O'Grady et al. 2006; Edmands 2007). The second is the loss of potentially adaptive genetic variation which limits the species' ability to adapt in response to environmental changes (Reed 2005; O'Grady et al. 2006; Wright et al. 2009; Tollington et al. 2013) such as, for example, climate change. The third is deleterious allele accumulation, also called "mutational meltdown," due to the fact that selection is weaker in small populations than in large populations (Reed 2005; O'Grady et al. 2006).

The last major risk for the Raso lark is increasing human pressure on the island. The population of Cape Verde has more than doubled in the past 50 years, and stood at 553,432 inhabitants in 2016 (CIA 2016). Tourism is also rising, with the number of tourists increasing from 24,000 in 1900 to 333,000 in 2008 (Ministry of Environment, Housing and Territory Planning of Cape Verde 2011). In addition to the direct negative impacts of people on Raso (nest or feeding hole trampling for example), each boat coming to the island increases the risk of predators being introduced. Donald and Brooke (2006) provide an excellent set of guidelines to help visitors to Raso reduce their impact.

These three potential threats, added to the species' extremely small range (one of the smallest in the world) and great fluctuations in population size, are the reasons why the Raso lark is classified as Critically Endangered according to criteria B1ac(iv)+2ac(iv) on the IUCN Red List (IUCN 2016).

As a result, the Raso lark has been protected by Cape Verdean law since 1995, and Raso is part of a National Park that also encompasses Santa Luzia and Branco islands since 1990 and for which access permission is needed (Donald and Brooke 2006). However, in practice, the Cape Verdean government lacks the financial and human resources to effectively guard, manage and protect the area. Since 2012, the Royal Society for the Protection of Birds (RSPB), the Sociedade Portuguesa para o Estudo das Aves (SPEA) and Biosfera I - respectively British, Portuguese and Cape Verdean conservation non-profits - have been undertaking a preliminary assessment of a

project to reintroduce the lark onto Santa Luzia (Figure 1.3). The habitat on Santa Luzia has already been analyzed and found suitable for the Raso lark, and the partners in the project are currently working on the next step: the eradication, or at least control, of cats on what may become the lark's second home.

Aim and subject of this thesis

My goal is to contribute to these efforts to protect the Raso lark, as well as to help gain a better understanding of island bird conservation in general. Of the main threats to the Raso lark detailed in the paragraphs above - alien predators, increased human presence on Raso, small range, climate change, extreme population fluctuations and potentially low levels of genetic diversity - the last two are the least well understood. Filling these gaps in our knowledge is the aim of the four original research chapters in this thesis. For this purpose, I used two types of data: a long-term capture-recapture dataset resulting from the ringing of Raso larks since 2002, and a genetic dataset produced by RAD sequencing. Both methodologies are presented and explained in Chapter 2. The capture-recapture dataset serves to investigate the causes of the extreme population fluctuations: Chapter 3 estimates year-to-year survival rates in adults, and Chapter 4 examines how the rates are affected by environmental parameters such as rainfall and individual parameters such as body size. The genetic dataset is used to compare the island endemic Raso lark with its two continental and widespread relatives, the skylark and the Oriental lark: Chapter 5 first resolves the phylogenetic node between the three species, and Chapter 6 then focuses on the Raso lark, estimating its levels of genetic diversity, studying its evolutionary history, and looking for specific genetic signatures that differentiate it from its relatives. Finally, to conclude, Chapter 7 uses this newly acquired information, in combination with previously existing knowledge, to inform conservation plans for the species, and in particular to make recommendations for the future reintroduction of the Raso lark onto Santa Luzia.

Chapter 2: General methods

Fieldwork, MARK and RADseq

A view from Raso towards the uninhabited islands of Branco (left) and Santa Luzia (right). © Elisa Dierickx

Abstract

This chapter gives a general overview of the three methodologies that are the basis of this thesis: fieldwork on Raso Island to gather data on the Raso lark (relevant to all chapters), the capture-recapture survival analysis software called MARK (chapters 3 and 4) and the genome fractioning sequencing method RADseq (chapters 5 and 6).

Fieldwork on Raso

Paul Donald visited Raso in October 2001 to conduct a census of the Raso lark (Donald et al. 2003). Further research progress depended on catching and marking birds. The feasibility of catching them was established during Dr. Michael Brooke's first visit in December 2002, when the first 57 birds were colour-ringed (Brooke and Bolton 2002). Since then, Dr. Brooke has returned to the island every single year for two-three weeks, during the months of November or December. I myself worked on Raso in 2013 and 2014.

Research conditions on Raso are not easy (Figure 2.1). Raso is a desert island, without inhabitants, buildings, electricity or water. Communication is very difficult. Cell phone reception is marginal, and only available on a few spots on the island. Access to the island is only possible with a fisherman's boat, and landing is rendered difficult by the surrounding cliffs. All food, water and equipment arrives on Raso on the first day; there are no fresh supplies during the next three weeks, since there is usually no contact with other people during that time, except for occasional visits by fishermen from nearby islands. These conditions and the high Cape Verdean temperatures dictated some of the methodological decisions that were made and which are detailed below.

Data was collected on Raso over the years with four main goals in mind: to estimate population size, to follow individual survival, to study morphological characteristics and to provide samples for different types of genetic analyses.

Figure 2.1 From top to bottom, the camp on Raso Island, the kitchen and the bathroom. © Elisa Dierickx

Boyd Alexander, who discovered the species, described Raso larks as "so tame that we could have knocked many over with sticks" (Alexander 1898). Today they are a little more wary of humans, but still reckless enough for the following capture methodology to succeed. Birds are caught using a mist net attached to poles, stretched out, held tight, and positioned horizontally above the ground by two people, one holding each pole. With the net held as described, birds are approached downwind, moving very slowly. Birds targeted are either unringed, or those missing one or more of their set of three colour rings. When a bird is sufficiently close, positioned towards the middle of the net and, ideally, facing it, it is caught by throwing the net over it, with the help of the strong winds that commonly occur on Raso. The bird is then quickly disentangled from the net. If processing of the bird cannot be done immediately, it is placed in a cotton bag, which is then hung in a shady place. Time and place of capture are written down.

The following measurements are made: tarsus length, wing length (flattened), tail length, bill length (both to feather and to skull) and bill depth and width. The bird is weighed using a spring scale. It is also checked for toe damage (scored as absent or present - and, if present, on which toe(s)), for absence or presence of a brood patch, and for absence or presence of moult of the remiges and rectrices. To minimize bias, all these measurements and observations were made by the same person, Michael Brooke, over the course of the study.

The bird is fitted with a metal ring inscribed with a unique ring number, and with a set of three Darvic plastic colour rings[1] (Figure 2.3). Each bird has a unique combination of colour rings, read from top to bottom, left leg before right leg. Available colours are white, black, blue, dark green, lime, yellow, orange and red. For example, the bird in Figure 2.3 has the combination "white, metal; white, orange."

Finally, a blood sample is taken from the bird's ulnar vein. The area is cleaned with ethanol and covered with a drop of EDTA. A small needle is then used to prick the vein so that the blood flows directly into the drop of EDTA. The liquid is then absorbed with a piece of filter paper marked with the bird's ring number. The bird is then released, and another drop of EDTA is added to the paper, which is then air-dried before being placed in an envelope and stored in silica gel. Upon return to the United Kingdom, all samples are stored at -80°C.

[1] Early experience in 2002 and 2003 with colour rings supplied by A.C. Hughes, the principal UK supplier, revealed that the lifetime of these rings in the Raso environment was very short - a mere 15 months.

During the entire time on Raso, great effort is made to read the colour-ring combinations of all encountered larks. The goal is to be confident at the end of the stay that virtually all colour-ringed individuals still alive have been sighted. Inevitably a few birds (< 10%) escape observation and are recorded alive the following year. Reading colour-ring combinations allows the fate of many individual birds to be followed over the years. It also yields an estimate of the total number of the colour-ringed birds present on the island on the annual visit.

This colour-ring information is critical for the estimation of total population size. The methodology for estimating the population size is as follows. On the last two days of the stay, when we are relatively certain that almost all living colour-ringed birds - both freshly-ringed birds and those ringed in previous years - have been sighted, two people each walk separate transects on each day (for a total of four transects). A transect consists of a random walk of a few hours, covering all the areas of the island where larks are likely to be encountered. On each transect, all encountered birds are counted, and classified as either “colour-ringed” or “unringed.” The ratio of colour-ringed to unringed birds can then be extrapolated to the whole population to yield a population size estimate. This estimate is then corrected the following year, to account for colour-ringed individuals that were spotted then but missed in the previous year.

In addition to these planned activities, nests are recorded opportunistically when encountered, often when a parent is flushed from the nest (Figure 2.2). Location, clutch size and progression of the clutch/brood are recorded, as well as the colour-ring combinations (if any) of the parents. When chicks are old enough, they are ringed and bled following the same protocol as adults (Figure 2.4).

Figure 2.2 Raso lark nests. © Elisa Dierickx

Figure 2.3 Newly colour-ringed Raso lark with the combination "white, metal; white, orange." © Elisa Dierickx

Figure 2.4 Newly-ringed Raso lark chick before return to its nest. © Elisa Dierickx

MARK

MARK (White and Burnham 1999) is a software application for fitting Cormack-Jolly-Seber (CJS) models to datasets with marked individuals (Lebreton et al. 1992). It allows us to calculate survival (Phi) and resighting (p) probability estimates based on animals that were marked and re-encountered again later on, either alive or dead, and either visually or through other devices such as radio tracking. To perform these calculations, MARK requires an input file with the encounter history for each individual. Time intervals between re-encounters can be uniform or not; this parameter was set to one year in my analysis. Animals can be divided into groups (for example, males and females, or treatment and control). Individual covariates (for example, individual wing length), group covariates (for example, mean female wing length) and independent covariates (for example, average rainfall for each year) can be considered in the analysis (Cooch and White 2014).

Goodness of Fit (GoF) testing to verify the assumptions underlying the models being fitted to the dataset is a crucial first step to ensure that the most general model in the candidate model set adequately fits the data. Indeed, comparing the relative fit of a general model with a reduced parameter model offers good inference only if the more general model satisfactorily fits the data (Amstrup, McDonald and Manly 2005; Cooch and White 2014).

CJS models make four assumptions (Lebreton et al. 1992). The first assumption is that all marked individuals in the population have the same chances of recapture within model strata at any time (i): all marked individuals should be equally "detectable" at occasion (i+1), independent of whether or not they were caught at occasion (i) (Lebreton et al. 1992). If females spend more time than males near or on the nest during breeding periods, they could potentially be more difficult to detect. However, this seems unlikely, based on preliminary year-to-year resighting calculations (Brooke and Dierickx 2013; Brooke and Dierickx 2014; Brooke and Mainwaring 2015), on experience in the field (birds tend to be sighted either when foraging in pairs, or, when conditions are not suitable for breeding, in large mixed-sex flocks), and the fact that there is not difference in the mean dates on which individual colour-ringed males and females are first seen (one would expect it to be later for females if they were more difficult to spot; Brooke and Flower 2011).

The second assumption is that, among the marked animals in the population, all have the same probability of surviving, regardless of when they were marked, within

model strata: individuals marked at occasion (*i*-1) have the same probability of surviving from (*i*) to (*i*+1) as do animals marked at occasion (*i*) (Lebreton et al. 1992). This is the case for the vast majority of my data points, except maybe for 31 birds that entered the dataset as juveniles and might, in their first year, have a lower probability of surviving than adults. However, this only concerns a very small proportion of the dataset, and both preliminary data and field observations suggest that juvenile survival in the Raso lark is very high (see Chapter 3).

The third assumption is that rings are not lost or missed (Lebreton et al. 1992). The data used for the analyses cover the time span 2004-2014. The colour rings used in 2002 and 2003 were not resistant enough for the Raso environment. For this reason, data on birds ringed during these two years could not be used for the MARK analyses. Their rapid loss would bias survival estimates downwards. Therefore birds first ringed in one of those years and given new colour rings in the years after 2003 were considered as having been first caught in the year that they were re-ringed. This affects 26 individuals.

The fourth assumption is that all sampling of animals is instantaneous and each release is made immediately after the sampling (Lebreton et al. 1992), which is the case for the Raso lark dataset.

GoF testing was performed for each dataset and the most complete CJS model (Lebreton et al. 1992). The two types of GoF testing used were the program RELEASE, which is integrated into MARK, and parametric bootstrapping. RELEASE uses two χ^2 contingency table analyses to test the first two CJS assumptions described above, which are commonly seen as the most problematic for capture-recapture datasets (Amstrup, McDonald and Manly 2005; Cooch and White 2014). Parametric bootstrapping uses simulation and resampling: based on the data, it first generates the distribution of model deviances, and then evaluates the observed value against this generated distribution (Amstrup, McDonald and Manly 2005; Cooch and White 2014). P-values larger than 0.05 were considered to indicate a satisfactory fit between the dataset and the model. $\hat{c}$ ("c-hat") - the variance inflation factor - was estimated for each dataset based both on RELEASE and parametric bootstrapping, and was adjusted in the analyses towards the highest value of the two (that is, towards the most conservative value) (Cooch and White 2014).

For each model tested, MARK calculates the quasi-likelihood adjusted Akaike's Information Criterion (AIC) as

$$AIC = \frac{-2\ln(L)}{\hat{C}} + 2K + \left(\frac{2K(K+1)}{n-K-1}\right)$$

where L is the model likelihood, $\hat{c}$ the variance inflation factor, K the number of parameters in the model, and n the sample size. $\hat{c}$ is a quantification of the over-dispersion of the data which adjusts AIC for the lack of fit between the data and models relative to a saturated model. Practically speaking, the "best" model is the model with the lowest AIC. This is the most parsimonious model given the data, the one with the best fit with fewest parameters (Burnham and Anderson 2002).

MARK ranks all the tested models from lowest AIC to highest AIC. It also calculates the difference in AIC between each model and the "best" model (ΔAIC). When ΔAIC between two models is < 2, then we can conclude that "both models have approximately equal weight in the data." If ΔAIC > 2, then there is evidence that the models are different (Anderson and Burnham 1999, Cooch and White 2014).

For each model, MARK also reports its deviance, defined as the difference in model likelihood −2 ln(L) between that model and the saturated model. Each model is also given a "normalized Akaike weight" (AIC weight), which indicates the likelihood of the model given all the other models tested (Akaike 1973). These weights (w_i) are calculated as follows for each candidate model (*i*) (Burnham and Anderson 2002, 2004):

$$w_i(AIC) = \frac{\exp\left\{-\frac{1}{2}\Delta_i(AIC)\right\}}{\sum_{K}\exp\left\{-\frac{1}{2}\Delta_K(AIC)\right\}}$$

To determine the relative importance of a variable (*v*) in explaining the data, the AIC weights of all the models in which that variable is found are summed ($\Sigma\ AIC_v$). This method is considered superior to classic significance testing for this type of analysis (Anderson and Burnham 1999, but see Cade 2015).

Models were then run with a logit link function by default, except when the program had difficulties estimating certain parameters. Indeed, sometimes MARK cannot estimate some of the model parameters based on the data (even though the parameters are, in principle, estimable). This usually manifests itself with estimates very close to 0 or 1 (and standard errors either equal to zero or very large), maybe because the data are too sparse, by chance unbalanced, or the parameter really is quite high (or low) (Cooch and White 2014). The logit link function is particularly sensitive to the latter. When MARK fails to estimate a certain number of parameters, the general

recommendation is that one should adjust the number of parameters from "estimated" to "estimable." Before doing that, it is advisable to use different methods in MARK to try to estimate as many parameters as possible, as described in the MARK user manual (Cooch and White 2014).

Finally, to explain the notation conventions used in this thesis and for these types of MARK analyses in general, consider a hypothetical example. A model including the effects of the covariates "time" (t), "sex" (s) and the interaction between these two, on both survival (Phi) and resighting (p), will be notated as follows: Phi(t*s) p(t*s), where "t*s" is shorthand for "t + s + t:s." A similar model *without* the interaction term will be notated as follows: Phi(s+t) p(s+t). Finally, the same model but with a constant value being estimated for resighting is written as: Phi(s+t) p(.).

RADseq

Restriction-site Associated DNA Sequencing (RADseq) is a fractional genome sequencing method consisting of the digestion of the genome with a restriction enzyme, followed by the attaching of adapters to the resulting DNA fragments. As a result, numerous Single Nucleotide Polymorphisms (SNPs) can be identified and analyzed. RADseq loci can be present anywhere in the genome - in both coding and non-coding areas. The cut sites are generally well conserved, so usually individuals from closely related species share most loci. The choice of the digestion enzyme (or enzymes in the case of double-digest RADseq) determines the size of the DNA fragments (Andrews et al. 2016).

The first step of RADseq is to fragment the target genome: each individual's DNA is cut with a restriction enzyme during a phase called "digestion." P1 adapters are then ligated to the restriction site end of the fragments in order to uniquely label the DNA of each sample with a specific barcode. This allows the pooling of DNA from all samples ("multiplexing"). Next the pooled DNA is randomly sheared through sonification. P2 adapters are then ligated to the other (sheared) end of the fragment. This is followed by an enrichment-PCR which aims at increasing the yield of RADseq. If the aforementioned steps are performed multiple times with different individuals each time, each batch (called a library) receives a different P2 primer during PCR, which allows for the pooling of all the libraries for Illumina high-throughput sequencing. This sequencing can be single-end or paired-end, have a higher or lower depth of coverage, and produce shorter or longer reads (Box 2.1; Figure 2.5).

After sequencing, the pooled samples are separated bioinformatically thanks to the P1 and P2 barcodes. The reads can either be aligned to a reference genome, or analyzed *de novo*, with software such as Stacks (Catchen et al. 2011), in which similar reads within each sample are grouped together, one on top of the other, forming "stacks." Each stack is first inspected column by column, to identify and record polymorphisms ("SNP calling"), and then row by row, to find each different haplotype (Catchen et al. 2011). Genotyping errors can also be found and corrected, by comparing read counts for each base at each position: real alleles will have more reads than errors (Davey and Baxter 2010).

This RADseq data can be used for population structure analyses, phylogenetics, the identification of SNPs, genome-wide association studies, and linkage and quantitative trait locus mapping, amongst others (Davey and Blaxter 2010). Its main advantages compared to other genotyping techniques are its non-reliance on prior knowledge of the DNA sequence, making it particularly useful for non-model species, its cheapness due to the fact that it only provides data on a small part of the genome, its rapidity, and the relatively large amount of data that it produces (tens of thousands of loci across the whole genome) (Davey and Blaxter 2010; Dierickx et al. 2015).

Box 2.1 Glossary of RADseq-related terms.

Glossary

Contig: a set of overlapping reads constituting a consensus region of DNA.

Depth of coverage: the number of times that a nucleotide is sequenced during a sequencing run. The higher the number of times that this nucleotide has been sequenced, the better the quality of the data, since this reduces the chance that a particular nucleotide or SNP is a sequencing error. Not all targeted nucleotides are sequenced an equal amount of times. Hence one can look at the average coverage per individual, population or sequencing run.

Haplotype: contraction of "haploid genotype." An individual's set of DNA variations (alleles or SNPs) found on a same chromosome and that tend to be inherited together.

Indel: an INsertion or a DELition of a base in a genome compared to another.

Primer: a short strand of DNA or RNA that is used as the starting point for DNA synthesis. Primers are necessary for PCR, because DNA polymerases can only add new nucleotides to an existing strand of DNA.

Read length: the number of base pairs that are read at a time by the sequencer (50, 100, or more). Longer reads are more expensive, but can help accurately determine the relative locations of specific base pairs.

Restriction enzyme: an enzyme that cuts DNA at a specific place that it recognizes. The enzyme "reads" the DNA molecule and recognizes a particular sequence, called the restriction site.

Single-end vs. paired-end read: when single-end reading is chosen, the sequencer reads the DNA fragments from only one end to the other to generate the sequence of base-pairs. In the case of paired-end reading, after this first step, it starts a second step of reading, this time from the opposite end. Paired-end reading makes it easier to position the various reads in the genome, which makes it the preferred method for addressing gene deletions, inversions and insertions for example. However, it is more expensive and time-consuming than single-end reading.

Scaffold: a group of contigs that is ordered and oriented. A scaffold has gaps, but there is normally evidence to support the contig order, orientation and gap size estimates.

SNP: a Single-Nucleotide Polymorphism is variation in a single DNA building block, called nucleotide. For example, most individuals might have the nucleotide thymine (T) at a specific position in the genome, but some might replace it with the nucleotide cytosine (C)

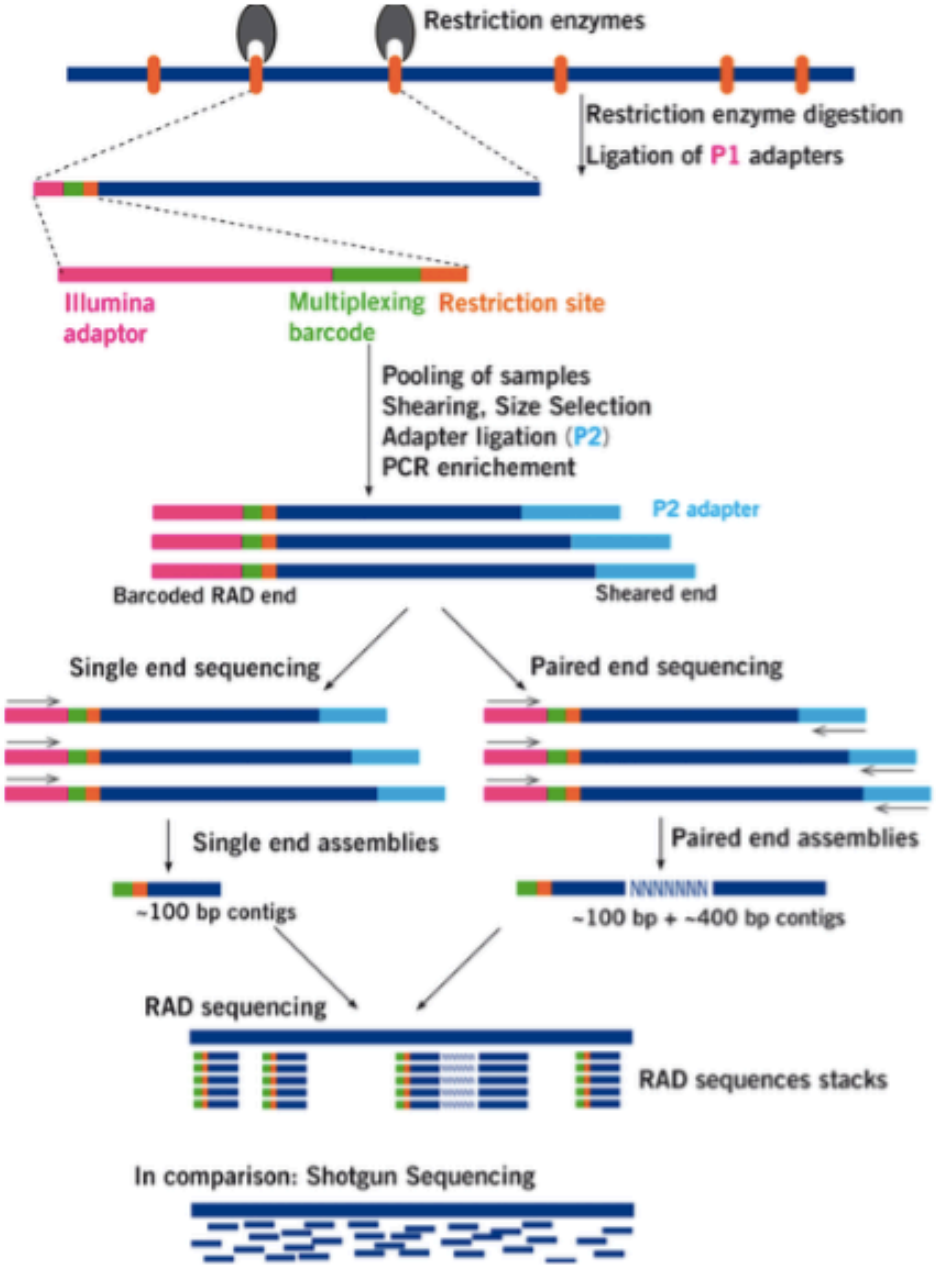


Figure 2.5 Overview of the RADseq method (figure retrieved on 07/12/2016 from http://www.floragenex.com/rad-seq/genome).

Chapter 3: Estimating adult survival in an island lark

Colour-ringed Raso lark male. © Edwin Winkel

Abstract

Estimating survival rates is crucial for the management of threatened species, including the Critically Endangered Raso lark. This chapter estimates adult survival in the Raso lark and tests whether it could be linked to two population phenomena observed in the field: a highly variable population size and a male-biased sex ratio in certain years. Using a capture-resighting dataset spanning 10 years, I estimated annual adult survival for both sexes to average 0.84 over this period. This is much higher than the survival rate of its closest relative, the skylark, but comparable to that of other island passerines. I also found strong evidence for survival varying among years (between 0.76 and 0.94) and differing between males and females (with males having higher survival until 2011, at which point the trend inverted), which could play a role in the aforementioned population size fluctuations and male-biased sex ratio in the study's early years, respectively. The next step in improving our ability to make informed conservation decisions will be to understand the factors behind these fluctuations in survival and sex ratio.

Introduction

Understanding the population dynamics of species is a fundamental component of evolutionary ecology, applied ecology and conservation biology. The goal is often to model population growth rates, to gain a better understanding of life history variation, and to apply this knowledge to conservation or wildlife management. The first step is to obtain reliable estimates of key demographic parameters, such as survival and reproductive output, through empirical studies based on field data (Monticelli et al. 2010). Estimating survival and understanding drivers of change in survival is key to understanding population changes and, hence, to informing conservation decisions (Siriwardena, Baillie and Wilson 2010). It can, for example, help conservationists identify particular environmental stresses on a species, such as climate change. Knowledge of survival rates, together with estimates of productivity, can also be used to identify the species' most limiting life stages on which focus is needed (Cox et al. 2014). For example, age-class based survival estimates for the Critically Endangered Maui parrotbill *Pseudonestor xanthophrys* showed that adult females had a survival rate significantly lower than adult males (Mounce et al. 2014). Adult female survival was also the most plastic rate in the study, that is, the rate with the most room for improvement. Based on this, the authors recommended striving towards decreasing adult female mortality, specifically through predator reduction, since the vulnerability of females on the nest was the most likely cause for the difference between male and female adult survival rates. A similar analysis on a red-backed shrike *Lanius collurio* population in the Netherlands led Hemerik et al. (2015) to conclude that juveniles and adult females were the best candidates for conservation efforts. They recommended increasing food availability on breeding grounds by creating a patchwork environment with a wide variety of ecological niches and prey.

In the case of the Critically Endangered Raso lark, whose entire population is confined to Raso Island, immigration and emigration do not need to be taken into account, and reproductive output and survival are the two most important vital rates left to estimate for this species. This chapter focuses on the latter: estimating survival in the Raso lark is its first aim. Studying survival at different life stages in the Raso lark is not possible at the moment because of limited data. However, we know that, out of 31 juveniles ringed over the course of the study, 71% survived until the following year. Out of 63 pulli ringed in 2009 and 2010, 63% survived to become adults (Brooke and Welbergen 2010; Brooke and Flower 2011). This suggests high juvenile survival

compared to other species (Campbell and Lack 1985; Newton 1998); in fact, it is comparable to the high *adult* survival rates found in other island passerines (Simon et al. 2001; Leech et al. 2007; Monticelli et al. 2010; Mounce et al. 2014).

In addition to conservation considerations, these survival rate calculations for the Raso lark will provide some empirical evidence concerning two theories about survival. The first focuses on the particularities of islands and the distinct impact that they could have on the survival rates of island species. Researchers frequently describe a pattern whereby island species and island populations of widespread species have higher survival rates than their continental counterparts (Monticelli et al. 2010; Covas 2012). The second (highly debated) conjecture is that tropical passerines have higher survival rates than their temperate counterparts (Johnston et al. 1997; Blake and Loiselle 2008, 2013). This conjecture is based on the fact that birds in the tropics generally lay smaller clutches than in the temperate zones, and on the inverse relationship between fecundity and survival (Johnston et al. 1997). As an island endemic and as a tropical species, this study on survival in the Raso lark may shed light on both of these hypotheses.

After estimating the adult survival rate in the Raso lark, the second aim of this chapter is to determine whether it significantly fluctuates over time. Indeed, the species' population size (N) has varied greatly, and higher adult mortality in certain years could contribute to the fluctuations. The population size has been estimated many times since 1965, including every year for the time period 2002-2016 (Table 1.1, Figure 3.1). Over the course of this most recent study period, population size was at its lowest in 2004 (N = 57) and reached a high point in 2011 (N = 1558).

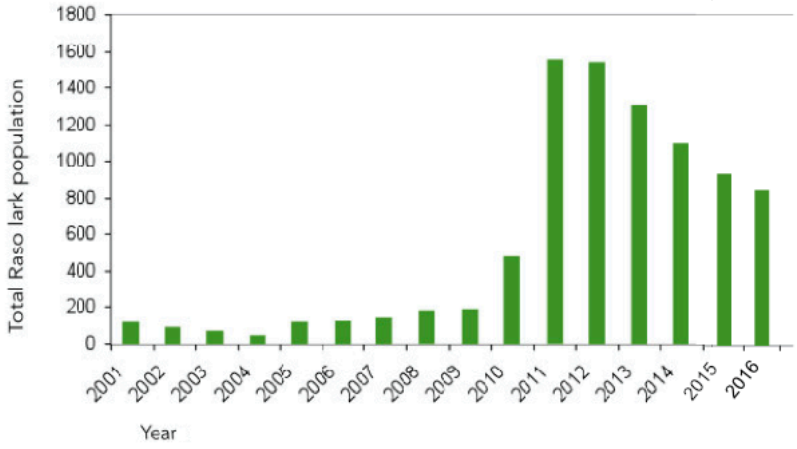


Figure 3.1 Total population sizes of the Raso lark (excluding nestlings) between 2001 and 2016 (Brooke and Gregory 2016).

Thirdly, there is evidence that, in certain years, the Raso lark's sex ratio is male-biased (Brooke and Bolton 2002; Brooke and Welbergen 2005), a phenomenon more common in endangered than in non-endangered species (Donald 2007). Could this bias in the Raso lark be due to males having a higher adult survival rate than females in certain years? In a study of 187 bird species from 59 families, Szekely et al. (2014) found that skewed sex ratios were linked to adult sex-biased mortality, but not to hatching sex or fledging sex ratios. This suggests that sex-biased adult mortality might be the cause of the skewed sex ratios frequently observed in birds (Szekely et al. 2014). This skew is indeed very common, and most often male-biased. Reporting on over 200 published avian studies, Donald (2007) found that, on average, males outnumbered females by around 33%, and that in 65% of cases the sex ratio was significantly biased. Similarly to Szekely et al. (2014), Donald (2007) also found that higher female mortality, as opposed to skewed offspring sex ratio, was the mechanism behind the sex ratio bias in birds.

However, while higher female mortality is often the reason behind male-biased sex ratios, it is far from being universal amongst birds. In fact, Jones et al. (2009) claim that males generally show lower survival than females. An example of this would be the work of Hemerik et al. (2015), which calculated a survival rate of 0.64 for older females and of 0.54 for older males in the red-backed shrike. A counter-example would be Mounce et al. (2014) who found a survival rate of 0.72 for adult females and of 0.82 for adult males in the Maui parrotbill. Common explanations for lower adult male survival are the higher energy and nutritional requirements of the larger sex and the costs of male-male competition (Searcy 1979; Promislow, Montgomerie and Martin 1992; Jones et al. 2009). Where lower adult female survival is recorded, it is usually thought to be caused by their larger reproductive costs (Hanssen et al. 2005; Descamps et al. 2009; Mounce et al. 2014, but see Santos and Nakagawa 2012). The complexities are summarised by Jones et al. (2009) who write that "it may be difficult to make predictions regarding vulnerability on the basis of size or sex across species (or even across taxonomic groups) that have very different life histories." Nevertheless, analysing survival differences between the sexes and their potential impact on the sex ratio *within* a species can contribute to understanding that particular species' life history, as well as to practical conservation measures.

In addition to being a conservation priority, the Raso lark is an excellent model species to study adult survival both because it touches on the various debates referred to above, and because the quality of our long-term dataset is very high. At the time of the

analyses in this chapter, the population had been followed for the past 13 years, during which approximately 600 birds have been ringed. Furthermore, the fact that the Raso lark is an island endemic - a closed population - allows much more precise survival estimates than are achieved in most (open) populations. Finally, many other parameters were recorded alongside individual resightings, data that will be utilized in Chapter 4.

Methods

The fieldwork to collect the data used in this chapter, the notation conventions and the methodology for the analyses - using the software MARK - are outlined in Chapter 2. The starting model in MARK was a full time (t) dependence model for survival (Phi) and resighting (p) that also included sex (s): Phi (s*t) p(s*t), with $\hat{c} = 1.16$ based on a bootstrap of 1000 replications. The p-value from a GoF test in RELEASE was 0.09, indicating that the model adequately fits the data. Sample sizes can be found in Table 3.1. Given the very high juvenile survival reported above, birds ringed as (non-pulli) juveniles were left in this dataset. Analyses were also performed a second time with these individuals only entering the dataset when first seen as adults, with no notable effect on the results.

Table 3.1 Male and female sample sizes, calculated as birds captured or resighted in each year of the study. Total population size is included in the table for comparison.

Year	Ringed males	Ringed females	All	Total population size
2004	15	11	26	57
2005	59	18	41	132
2006	57	27	84	140
2007	58	41	99	159
2008	71	42	113	184
2009	65	47	112	193
2010	98	76	174	486
2011	158	140	298	1558
2012	186	168	354	1546
2013	175	149	324	1314
2014	137	148	285	1101

Results

All models were ranked based on their AIC weights - models with higher values carry more weight with the data, and therefore rank higher. Three models have ΔAICs below 2, and hence carry similar weight with the data and cannot be differentiated: Phi(s*t) p(.), Phi(t) p(.) and Phi(s*t) p(s). There is very little support for the other models, with ΔAICs larger than 2 and AIC weights close or equal to 0 (Table 3.2). Hence we can reject them. Adding up the AIC weights of all the models in which each variable is found, I obtained $\Sigma\ AIC_{time} = 1$ for survival, $\Sigma\ AIC_{time} = 0.01$ for resighting, $\Sigma\ AIC_{sex} = 0.75$ for survival and $\Sigma\ AIC_{sex} = 0.29$ for resighting.

Table 3.2 All models tested, including the fully sex and time dependent model (†) and the fully constant model (††).

Model	ΔAIC	AIC weight	No. of parameters	Deviance
Phi(s*t) p(.)	0.0	0.36	21	303.75
Phi(t) p(.)	0.7	0.25	11	324.90
Phi(s*t) p(s)	0.9	0.23	22	302.58
Phi(s+t) p(.)	2.7	0.09	12	324.82
Phi(s+t) p(s)	4.0	0.05	13	324.07
Phi(s*t) p(s+t)	5.0	0.03	29	292.26
Phi(s*t) p(t)3	8.1	0.01	29	295.38
Phi(t) p(s*t)	8.4	0.01	26	301.85
Phi(t) p(s+t)	10.2	0.00	21	313.90
Phi(s*t) p(s*t)†	17.4	0.00	40	281.65
Phi(.) p(.)††	29.4	0.00	2	371.70
Phi(.) p(t)	31.1	0.00	11	355.26
Phi(s) p(.)	31.4	0.00	3	371.70
Phi(.) p(s*t)	36.6	0.00	21	340.36
Phi(s) p(s*t)	38.1	0.00	22	339.81
Phi(s) p(s+t)	46.7	0.00	13	366.76

Survival estimates based on the most supported model, Phi(s*t) p(.), vary between 1 (SE = 0.0) for the time period 2007-2008 and 0.71 (SE = 0.07) for the time period 2012-2013 in males. For females, they range from 0.56 (SE = 0.17) for the time

period 2004-2005 and 0.94 (SE = 0.06) for the time period 2006-2007. In earlier years, male survival appears to have been higher than female survival, but the trend reverses from 2010-2011. On average, these trends seem to cancel each other out: over the whole time period 2004-2014, the average difference in survival between the sexes ≈ 0 (SE = 0.01), as estimated by model Phi(s) p(.). The largest differences in survival between males and females were 0.44 in 2004-2005, 0.13 in 2005-2006 and 0.16 in 2007-2008. The 95% confidence intervals are larger in the earlier years of the study when sample sizes were smaller (Figure 3.2). Still based on model Phi(s*t) p(.), the re-sighting rate (p) of 0.90 (SE = 0.01) does not vary significantly over time, and is similar for males and females.

The second most supported model, Phi(t) p(.) estimates similar survival rates for both sexes and ranging from 0.94 (SE = 0.03) in 2007-2008 to 0.76 (SE = 0.05) in 2012-2013, averaging 0.84 (SE = 0.01). The resighting rate is the same as for model Phi(s*t) p(.). The third most supported model, Phi(s*t) p(s), estimates the same survival rates as model Phi(s*t) p(.), but indicates a small difference in re-sighting rates between the sexes: 0.91 (SE = 0.01) for males and 0.89 (SE = 0.02) for females.

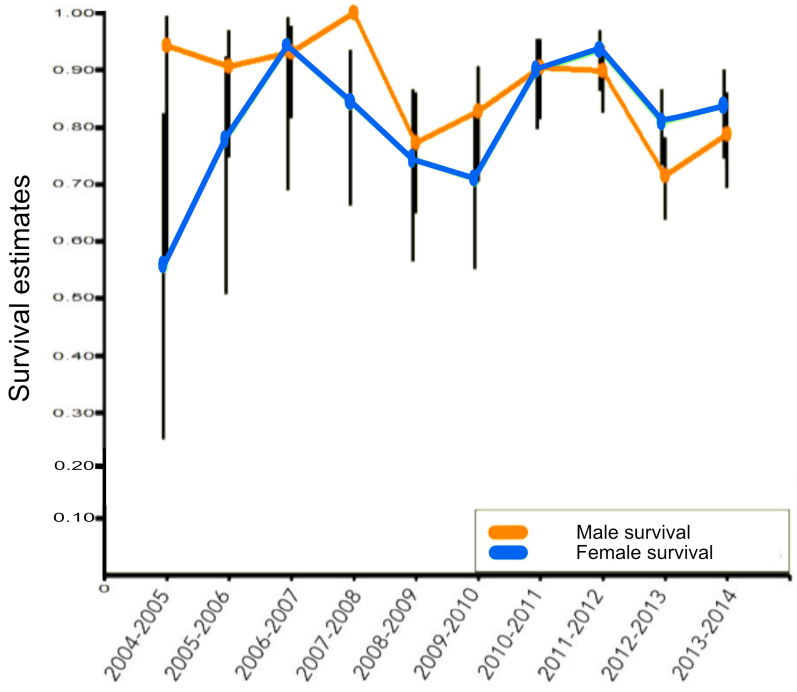


Figure 3.2 Survival estimates for males (in orange) and females (in blue) for the time period 2004-2014, based on the model Phi(s*t) p(.). 95% confidence intervals are indicated by the black error bars. As explained in Chapter 2, these can sometimes be implausibly large when survival approaches 1, due to MARK's difficulty in calculating such high survival estimates.

Discussion

I conducted a MARK analysis to estimate the survival and resighting rates of adult Raso larks. I used a capture-resighting dataset of colour-ringed individuals observed from 2004 until 2014. For this time period, I found a generally high survival rate, fluctuating between 0.76 and 0.94, with an average of 0.84. There was overwhelming support for a year effect on survival, that is, there is strong evidence for survival changes over time (Σ AIC $_{time}$ = 1.00). The survival estimates can be used to calculate life expectancy in different years, using the formula

$$E = B_a - \frac{1}{\ln(Phi)}$$

where E stands for life expectancy, B_a for age at first breeding and Phi for survival probability. Based on dramatic past annual population increases, age at first breeding in the Raso lark may be as young as 5 months (Michael Brooke, personal communication). Hence, life expectancy fluctuates between 4.1 and 16.6 years depending on the year and assuming no changes in survival with age.

The Raso lark's annual adult survival rate is high compared to the majority of passerine bird species. Its closest relative, the skylark, has a much lower annual survival rate, estimated by different researchers between 0.39 and 0.78, with most studies placing it around 0.50-0.60 (Donald 2004). The Raso lark's survival rate is also much higher than that of most other continental passerines. Blake and Loiselle (2008) estimate average survival rates of forest species in Eastern Ecuador at 0.58. Another survey found a mean survival rate of 0.53 in North American passerines and of 0.65 in Trinidadian passerines (Johnston et al. 1997). Other island passerines, however, seem to generally have higher survival rates, comparable to that of the Raso lark (Simon et al. 2001; Leech et al. 2007; Monticelli et al. 2010; Mounce et al. 2014). It is worth noting that calculating reliable "true" survival estimates can be extremely challenging, which may explain for example the wide range of estimates for the skylark. Many studies struggle to maintain a marked population for long enough and are limited by the important unknown that is emigration, which skews survival estimates down.

I found strong evidence for a difference in survival between males and females (Σ AIC $_{sex}$ = 0.75). Two of the three most supported models include an interaction term between time and sex, suggesting that the influence of sex on adult survival varies over time. Indeed, while males seem to have had higher survival at the beginning of the study, this trend has been reversed since 2011-2012 (Figure 3.2). This resulted in no

difference in survival between males and females on average for the whole 2004-2014 period, potentially explaining why it was hard to distinguish between the Phi(s*t) p(.) and Phi(t) p(.) models. Field observations confirm this reversal of trends in survival between males and females in recent years, and females have recently made up a majority of newly-ringed birds (Brooke and Dierickx 2013; Brooke and Dierickx 2014; Brooke and Mainwaring 2015), which stands in stark contrast with the male-biased sex ratio observed in the past. Since in certain years (disregarding the 2004-2005 time period with higher uncertainty in the estimates) males have a survival rate that is up to 16% higher than females, this difference could lead to the male-biased sex ratio sometimes observed in the population, especially if lower female survival happens to coincide with years of little breeding, which would tend to reduce the bias by injecting equal numbers of males and females into the population (given an equal primary sex ratio and early survival).

This study also investigated resighting rates, and found them to be high and constant over time. There is a small possibility that resighting rates are different for males and females (Σ AIC $_{sex}$ = 0.29). However, even if this was the case, the difference in resighting for males and females would be very small - 0.91 and 0.89 respectively. This indicates that the sex ratio estimates, and hence the suggestion of the strong male bias in the early years of the study, are likely to be correct.

In conclusion, this research supports the reality of the male-biased sex ratio observed in the early years of the study and suggests a mechanism for the generation of biases. It also estimated the adult survival rate of the Raso lark and found it to be very high compared to its sister species, the continental, temperate skylark, and to passerines in general. It is, however, comparable to that of other island-dwelling birds. This supports the possibility that island species and tropical passerines have higher survival rates than their continental and temperate counterparts, respectively. Finally, this research showed that the survival rate of the Raso lark significantly varied over time, as well as between males and females - two phenomena of which the causes will be studied in Chapter 4. Chapter 7, in turn, will discuss the practical implications of these results for the conservation of the Raso lark.

Chapter 4: What influences adult survival in the Raso lark?

Colour-ringed Raso lark male. © Edwin Winkel

Abstract

From both a conceptual and a nature conservation perspective, there is considerable interest in developing theoretical frameworks to predict survival. Some of the existing frameworks, based on environmental, population or individual characteristics, are heavily debated. Using MARK on a capture-recapture dataset spanning 10 years, this chapter aims at understanding what factors shape survival in the Critically Endangered Raso lark. Two types of variables were considered as potential impact factors: year-dependent (rainfall, population size, population mean clutch size) and individual-dependent (age, body size characters, size ratio with mate, Ase18 genotype). Amongst the year-dependent variables, only same-year rainfall impacted survival, with a 13% decrease in survival in the wettest year (2007) compared to the driest year (2010), making it the most likely explanation for the inter-annual fluctuations in survival found in Chapter 3. Results also hint at some of the individual factors - namely morphological measurements and Ase18 genotype - influencing survival. Overall, the picture that emerges from Chapters 3 and 4 is that of a species whose life history strategy is to invest heavily in maintenance and survival, but less into fecundity, which stands in sharp contrast with its closest relative, the mainland-dwelling skylark. This provides

evidence for the theory that island bird species generally have slower life history strategies than their continental counterparts.

Introduction

Chapter 3 showed that the survival rate of the Raso lark fluctuated significantly over the years, and that these fluctuations differed between males and females. What factors could explain these variations and differences? Understanding the parameters influencing survival rates is crucial for the management of endangered species, including the Critically Endangered Raso lark. From both a practical and a conceptual point of view, there is considerable interest in developing theoretical frameworks to predict survival. We can use our very complete long-term capture-recapture dataset, which also includes morphological measurements and blood samples, to improve our understanding of the factors influencing the Raso lark's survival and to situate it within the scientific literature. This topic has a number of sub-questions, hypotheses and predictions, detailed below.

Year-dependent variables

A first approach is to investigate the link between environmental, year-dependent factors and survival. In the case of the Raso lark, rainfall is the most obvious candidate, since it could be a strong limiting factor on arid Raso Island, and population change has been shown to be positively correlated with rainfall (Brooke et al. 2012; Figure 4.1). This is the case for several Darwin's finches species in the genus *Geospiza* on the Galapagos, where low rainfall negatively impacts individual survival due to reduced food availability (Grant and Boag 1979). In the Raso lark, is adult survival lower when rainfall is low? Does low rainfall affect females and males differently? My prediction is that rainfall is positively correlated with survival for both sexes, and that this effect is more pronounced for females than for males, as authors have previously suggested that males outcompete females for food in times when it is scarce (Donald et al. 2007).

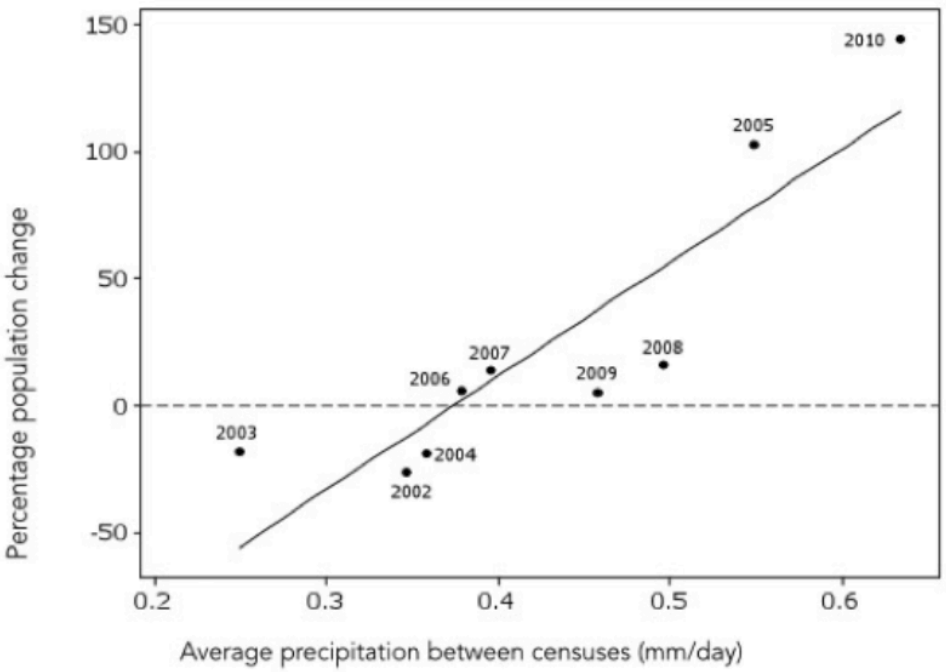


Figure 4.1 Correlation between change in population size and average precipitation between Raso lark censuses (copied from Brooke et al. 2012).

Another year-dependent candidate parameter is population size. Could survival of individuals be constrained by intra-specific competition? A decrease in survival, reproductive output, or both is expected in populations approaching their demographic equilibrium, through a negative feedback loop (Tavecchia et al. 2007; Monticelli et al. 2010), as shown for example in a population of Audouin's gulls *Larus audouinii* (Tavecchia et al. 2007). Since resources may be scarce on this desert island, my hypothesis is that in the Raso lark survival and population size are similarly negatively correlated.

The third, and last, year-dependent parameter examined in this chapter is mean clutch size. Does survival in one year correlate with breeding output in the same year? One hypothesis is that breeding takes a toll on the parents, and that adult survival, especially in females, the sex that incurs the high incubation costs (Hanssen et al. 2005), is low when mean clutch size is high. An alternative hypothesis is that, on the contrary, survival is positively correlated with mean clutch size, since potentially birds invest more in breeding in years with good environmental conditions and abundant resources allowing high survival. A third hypothesis is that survival does not correlate with mean clutch size, since the larks could adjust the level of their investment in breeding quite precisely to environmental conditions, and maintain constant survival.

Individual-dependent variables

Another approach is to focus on the characteristics of individuals, such as body size (e.g. Jones et al. 2009), song quality (e.g. Lambrecht et al. 1985), sex (e.g. Jones et al. 2009), genotype (e.g. Worley et al. 2010) or age (e.g. Hernandez-Matias 2011). The "best" song, size, sex or age can vary from species to species, from population to population and from year to year. There have been many attempts to understand the mechanisms behind this variation and untangle the respective effects of each individual variable on survival (for example, effects of larger body size can be confounded by the effects of sex, since, in birds, males are usually the larger sex).

In the Raso lark, the first individual factor that I investigated is age. Does adult survival decline with age? The oldest birds in our dataset were ringed as adults in 2004. Ten years later, some of these birds were still alive, as well as some potentially even older birds ringed in 2002 but not included in these analyses. Even though the Raso lark is long-lived for a passerine (see Chapter 3), I expect individuals to show signs of senescence in the later years of this study, as indicated by a decrease in survival as birds grow older.

Next, I investigated whether adult survival correlates with size. If yes, does the pattern vary according to sex and time? One hypothesis, amongst others, is that larger females and males are better at outcompeting their conspecifics for territories and resources so that larger individuals of both sexes have higher survival, and that this effect is more pronounced in years with scarce resources. This has been previously suggested by Donald et al. (2007) based on behavioural field observations.

An alternative hypothesis takes into account that the species is sexually dimorphic (males are larger than females) (Donald et al. 2003) and that individuals are often observed foraging in pairs: this dimorphism could have developed to avoid intersexual competition (Schülke and Kappeler 2003; Kamilar and Pokempner 2008) and, as a result, one would expect size and survival to be negatively correlated in females, and positively correlated in males. Similarly to the first hypothesis, this effect is expected to be more pronounced in years with scarce resources. To explore this alternative hypothesis, I investigate whether birds in pairs where the male/female size ratio is greater have higher survival.

Finally, I made use of the fact that a number of Raso larks were genotyped for different microsatellite loci for a previous study (Brooke et al. 2010; Table 4.S1). In small populations such as the Raso lark, diversity can be maintained through

mechanisms such as balancing natural selection, including heterozygote advantage (Hedrick 2007; Brooke et al. 2010). As a way of testing this, numerous studies investigated the link between survival and a particular genotype (e.g. Cohas et al. 2009; Christiansen et al. 2010; Worley et al. 2010; Canal, Serrano and Potti 2014). My goal is to investigate the links between survival and genotype at one known Raso lark locus of particular interest. In their study on genetic variability in the Raso lark, Brooke et al. (2010) found that, for one microsatellite locus, Ase 18, males could be either homozygous or heterozygous while females were always heterozygous, and always carried the female-only allele, 205, presumably on the W sex chromosome, and another of six possible alleles on the Z chromosome (Table 4.1). Thus this locus is sex-linked in the Raso lark, although not in the species for which the microsatellite primers were developed. Brooke et al. (2010) suggested that, since locus Ase18 was the only one of the 21 surveyed to have more than two alleles, it might be located next to a gene under selection for variation, making it particularly interesting for my study.

Table 4.1 Number of times each Ase18 allele was sequenced in males, in females and in total, from birds sampled between 2004 and 2010 (Brooke et al. 2010; Brooke, Komdeur and van der Velde, personal communication; Table 4.S1).

Allele name	Male count	Female count	Total
203	158	64	222
205	0	129	129
214	3	1	4
218	55	14	69
220	7	7	14
222	1	1	2
224	102	42	144

Methods

The fieldwork to collect the data used in this chapter, the notation conventions and the methodology for the analyses - using the software MARK - are outlined in Chapter 2. When defining models, the number of year-dependent variables was kept low, in order to respect the limitations of the 10-year longitudinal dataset. Sample sizes can be found in Table 3.1. The starting model was the one that was most supported in

Chapter 3: Phi (s*t) p(.). Since neither RELEASE nor bootstrapping are feasible in MARK for individual covariates, I followed the recommendation to use the same ĉ adjustment for these analyses as for the most general time-dependent model (ĉ = 1.16) (Cooch and White 2014). The more specific methods for each parameter are detailed below.

Rainfall

The rainfall data were downloaded from the website http://climexp.knmi.nl. More specifically, it corresponds to a satellite-based monthly Tropical Rainfall Measuring Mission (TRMM) precipitation index centred on Raso at 14°-19° N, 22°-27 W. Units are mm/day. These data differ from the data used by Brooke et al. (2012), who also investigated the effects of rainfall on the Raso lark, and used values for rainfall that combine data from the TRMM and from the Global Precipitation Climatology Centre (GPCC). I was not able to do so because the data from GPCC have not been available since 2013. Apart from this, in an effort to build a solid basis for comparison, I used the same methodology as Brooke et al. (2012) and created two estimates of annual rainfall, named "rain1" and "rain2." "Rain1" corresponds to rain with a delayed effect on survival. When estimating survival until year Y, "rain1" is the mean rainfall over the period spanning from two months before the census month in year Y − 1 until (and including) two months before the census month in year Y. "Rain2" corresponds to rain with an immediate effect on survival. When estimating survival until the year Y, "rain2" is the mean rainfall over the period spanning from a month after the census month in year Y − 1 until (and including) the census month in year Y. Although these two measures only differ by four months, they contain the rainy seasons of two different years, since rainfall is seasonal in Cape Verde, and they are not correlated (r = − 0.1, N=10).

Population size

Population sizes were estimated every year following the methods described in Chapter 2, and can be found in Table 1.1 and Figure 3.1.

Mean clutch size

Raso lark nests were discovered opportunistically, usually when a nesting bird was flushed from eggs or very small chicks. When a nest was discovered, the number of eggs or chicks was recorded. The mean clutch size for each year was calculated (Table 4.2).

Table 4.2 Mean clutch size and number of nests found for each year in the study.

Year	Mean clutch size	Number of nests
2004	2.04	11
2005	0.00	0
2006	2.00	4
2007	1.00	1
2008	1.54	17
2009	3.10	8
2010	3.10	29
2011	1.39	31
2012	1.00	5
2013	1.17	41

Age

Since most birds in our dataset were ringed as adults, I could not determine each individual's exact age. As a result, I used "time since first marking" as a proxy for age. The sample sizes were the same as described in Table 3.1.

Size

The dataset is the same as the one presented in Table 3.1, except that 31 birds (16 females and 15 males) measured as juveniles, as well as two females of uncertain age, were excluded from this analysis, since they potentially might have not been full-grown at the time of measurement. Measurements were taken in the field as described in Chapter 2. Three morphological measurements were used to reflect size: bill length

from tip to feathers (b), tarsus length (tars) and flattened wing length (w). All three measurements were analyzed separately, since specific selective pressures could be acting on bill size given the Raso lark's digging behaviour for feeding (see Chapter 1), and a Principal Component Analysis would not be suitable for the only two remaining measurements. Males and females were analyzed separately to avoid serious multicollinearity, since body size and sex are highly correlated in the Raso lark (there is little to no overlap in size between the sexes). The starting model for this analysis was Phi(t) p(.).

Size ratio with mate

For 100 mated birds (50 pairs), I calculated the bill length ratio (B), tarsus length ratio (TARS) and wing length ratio (W) between male and female. For example, $W = \frac{w_{male}}{w_{female}}$. A mated pair was defined as a pair breeding together, tending a nest with eggs or chicks. Survival of birds ringed before the year that they were seen in a pair was not taken into account. The year that the birds were first recorded in a breeding pair counted as the starting point. Six birds were observed with a different, second mate over the years. There is strong evidence that these changes happened after the death of the previous mate (Michael Brooke, personal communication). For each of these six birds, a pair was randomly selected for the dataset and the other excluded. The starting model for this analysis was Phi(s*t) p(.).

Genotype

I performed nine different MARK analyses (named A-I) to investigate any link between survival and Ase18 genotype (Table 4.3). Analysis A investigates whether Ase18 homozygous males have lower survival than heterozygous males (not differentiating between allele combinations), which could be the case according to heterozygote advantage theory. Analyses B and C research whether the trends observed in A are driven by the most common allele, allele 203. Analysis D asks whether having even a single 203 allele could confer higher survival to males with that allele. Analysis E investigates whether having two copies of allele 203 is more advantageous than having a single 203 copy. Analysis F checks whether 203 is more advantageous in certain years, and 224 in others, hence maintaining them both in large proportions in the

population. Analyses G, H and I asks whether being a 203 heterozygote confers survival advantages compared to other heterozygote allele combinations in females.

The sample sizes for each group in the different analyses of genotype and survival can be found in Table 4.4. The starting model was Phi(genotype*t) p(genotype*t).

Table 4.3 Description of the different MARK analyses conducted to investigate the link between Ase18 genotype and survival. Group 1, Group 2 and Group 3 (only in analysis F) are the groups whose survival rates are being compared in each analysis. The groups differ in their Ase18 genotype.

Analysis name	Group 1	Group 2	Group 3/Notes
A	All homozygous males	All heterozygous males	
B	Males homozygous for allele 203	All heterozygous males	
C	Males homozygous for alleles 218 and 224	All heterozygous males	
D	Males heterozygous with allele 203	All heterozygous males without allele 203	
E	Males homozygous for allele 203	Males heterozygous with allele 203	
F	Males homozygous or heterozygous with allele 203, without allele 224	Males homozygous or heterozygous with allele 224, without allele 203	Group 3 included in the analysis: males heterozygous with alleles 203 and 224
G	Females heterozygous with allele 203	All heterozygous females without allele 203	
H	Females heterozygous with allele 203	Females heterozygous with allele 224	Allele 224 is the third most common allele, after 205 (which all females have) and 203
I	Females heterozygous with allele 203	Females heterozygous with neither allele 203 nor allele 224	

Table 4.4 Sample sizes for each group across all years in the different genotype analyses.

Genotype analysis	ĉ	Sample size Group 1	Sample size Group 2	Sample size Group 3
A	1.13	45	106	N/A
B	1.10	34	106	N/A
C	1.15	12	106	N/A
D	1.15	25	68	N/A
E	1.15	34	24	N/A
F	1.18	56	35	53
G	1.08	64	65	N/A
H	1.14	64	42	N/A
I	1.18	64	23	N/A

Results

Year-dependent parameters

All models were ranked based on their AIC weights - models with higher values carry more weight with the data and rank higher. The most supported model was Phi(rain2) p(.), that is, a model containing rainfall with an "immediate effect" on survival as a variable. This was the only year-dependent variable that had an effect on survival, with Σ AIC = 0.74 (Table 4.5). Therefore while there was evidence for an effect of "immediate" rainfall (rain2), there was none for "delayed" rainfall (rain1), population size or breeding conditions as evidenced by clutch size, on survival. Annual survival is lower in years with high levels of rainfall, with a difference of 13% between the wettest year (2007) and the driest year (2010) (Figure 4.2).

Table 4.5 All models tested in the year-dependent variables analysis, as well as the fully constant model (†) for comparison.

Model	ΔAIC	AIC weight	No. of parameters	Deviance
Phi(rain2) p(.)	0.0	0.43	3	358.81
Phi(s+rain2) p(.)	2.0	0.16	4	358.76
Phi(s*rain2) p(.)	2.2	0.15	5	356.98
Phi(s*N) p(.)	2.2	0.14	5	357.00
Phi(s*rain1) p(.)	4.1	0.06	5	358.90
Phi(rain1) p(.)	4.8	0.04	3	363.62
Phi(s+rain1) p(.)	6.8	0.01	4	363.62
Phi(.) p(.)†	10.9	0.00	2	371.70
Phi(N) p(.)	11.3	0.00	3	370.07
Phi(B) p(.)	11.7	0.00	3	370.51
Phi(s) p(.)	12.9	0.00	3	371.70
Phi(s+N) p(.)	13.2	0.00	4	370.03
Phi(s+B) p(.)	13.7	0.00	4	370.50
Phi(s*B) p(.)	14.6	0.00	5	369.38

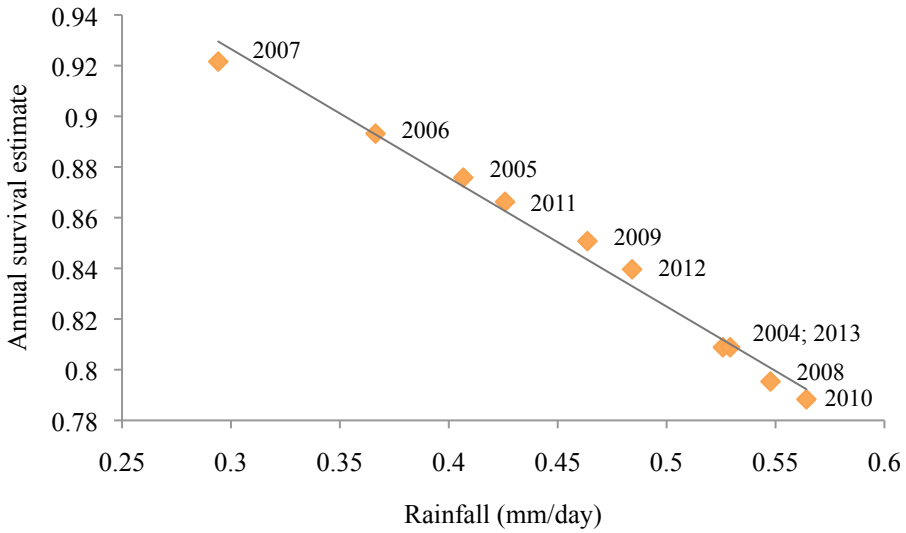


Figure 4.2 Annual survival estimates based on model Phi(rain2) p(.) in relation to mean rainfall in the same year.

Individual-dependent parameters

The age (or "time since first marking") variable had no significant explanatory power for either survival or resighting. The models including that variable had Σ AIC ≈ 0 for both survival and resighting. Therefore there was no evidence for an effect of age on survival.

For both females and males, for all three morphological measurements, the model Phi(t) p(.) ranks the highest. However, for all these analyses, except the female wing length analysis, a model containing the body size variable also has a ΔAIC value below or equal to 2.0, carrying similar weight with the data, and cannot be differentiated from the Phi(t) p(.) model. There is very little evidence for the other models, with ΔAICs larger than 2 and AIC weights close or equal to 0 (Table 4.6). Hence we reject them confidently.

Adding up the AIC weights of all the models in which each variable is found, I obtained, for the survival of females, $\Sigma\ AIC_{time\ (average\ of\ the\ 3\ measurements)} = 0.59$, $\Sigma\ AIC_{bill} = 0.43$, $\Sigma\ AIC_{tarsus} = 0.29$, and $\Sigma\ AIC_{wing} = 0.27$. For the survival of males, I obtained $\Sigma\ AIC_{time\ (average\ of\ the\ 3\ measurements)} = 1.00$, $\Sigma\ AIC_{bill} = 0.29$, $\Sigma\ AIC_{tarsus} = 0.27$, and $\Sigma\ AIC_{wing} = 0.47$. As for the direction of the effect, based on the additive models, it appears that larger body size, as approximated by bill and tarsus length, is associated with lower survival in females, a difference of ≈ 10% between the largest and the smallest females (Figure 4.3 A, C). Wing length has the opposite effect on female survival, but the model showing the trend is not supported by the data (Figure 4.3 E, Table 4.6). For males, the results are less clear-cut: bill length is positively correlated with survival, with a difference of ≈ 4% between the largest and the smallest bill length (Figure 4.3 B). Wing length is negatively correlated with survival, with a difference of ≈ 9% between the largest and the smallest wing length (Figure 4.3 F). Tarsus length has both little effect and little support (Figure 4.3 D, Table 4.6) over the 2004-2014 time period.

Table 4.6 Models with an AIC weight ≥ 0.00001 in the body size variables analysis, as well as the fully constant model, are included in this table.

	Model	ΔAIC	AIC weight	No. of parameters	Deviance
FEMALES					
Bill	Phi(t) p(.)	0.0	0.32	11	746.6
	Phi(.) p(.)	0.5	0.25	2	765.5
	Phi(b) p(.)	0.7	0.22	3	763.7
	Phi(t+b) p(.)	0.9	0.21	12	745.5
	Phi(t*b) p(.)	16.7	0.00	21	742.3
Tarsus	Phi(t) p(.)	0.0	0.40	11	746.6
	Phi(.) p(.)	0.5	0.31	2	765.5
	Phi(t+tar) p(.)	1.6	0.17	12	746.2
	Phi(tar) p(.)	2.4	0.12	3	765.4
	Phi(t*tar) p(.)	14.9	0.00	21	740.5
Wing	Phi(t) p(.)	0.0	0.41	11	746.6
	Phi(.) p(.)	0.5	0.32	2	765.5
	Phi(t+w) p(.)	2.1	0.15	12	746.6
	Phi(w) p(.)	2.5	0.12	3	765.5
	Phi(t*w) p(.)	11.0	0.00	21	736.7
MALES					
Bill	Phi(t) p(.)	0.0	0.71	11	838.3
	Phi(t+b) p(.)	1.8	0.29	3	838.1
	Phi(t*b) p(.)	12.3	0.00	12	829.8
	Phi(.) p(.)	20.5	0.00	2	877.3
	Phi(b) p(.)	22.4	0.00	3	877.0
Tarsus	Phi(t) p(.)	0.0	0.73	11	838.3
	Phi(t+tar) p(.)	2.0	0.27	12	838.3
	Phi(t*tar) p(.)	15.0	0.00	21	832.5
	Phi(.) p(.)	20.5	0.00	2	877.1
	Phi(tar) p(.)	22.4	0.00	3	877.0
Wing					
	Phi(t) p(.)	0.0	0.53	11	838.3
	Phi(t+w) p(.)	0.2	0.47	12	836.5
	Phi(t*w) p(.)	15.5	0.00	21	833.0
	Phi(.) p(.)	20.5	0.00	2	877.1
	Phi(w) p(.)	21.7	0.00	3	876.3

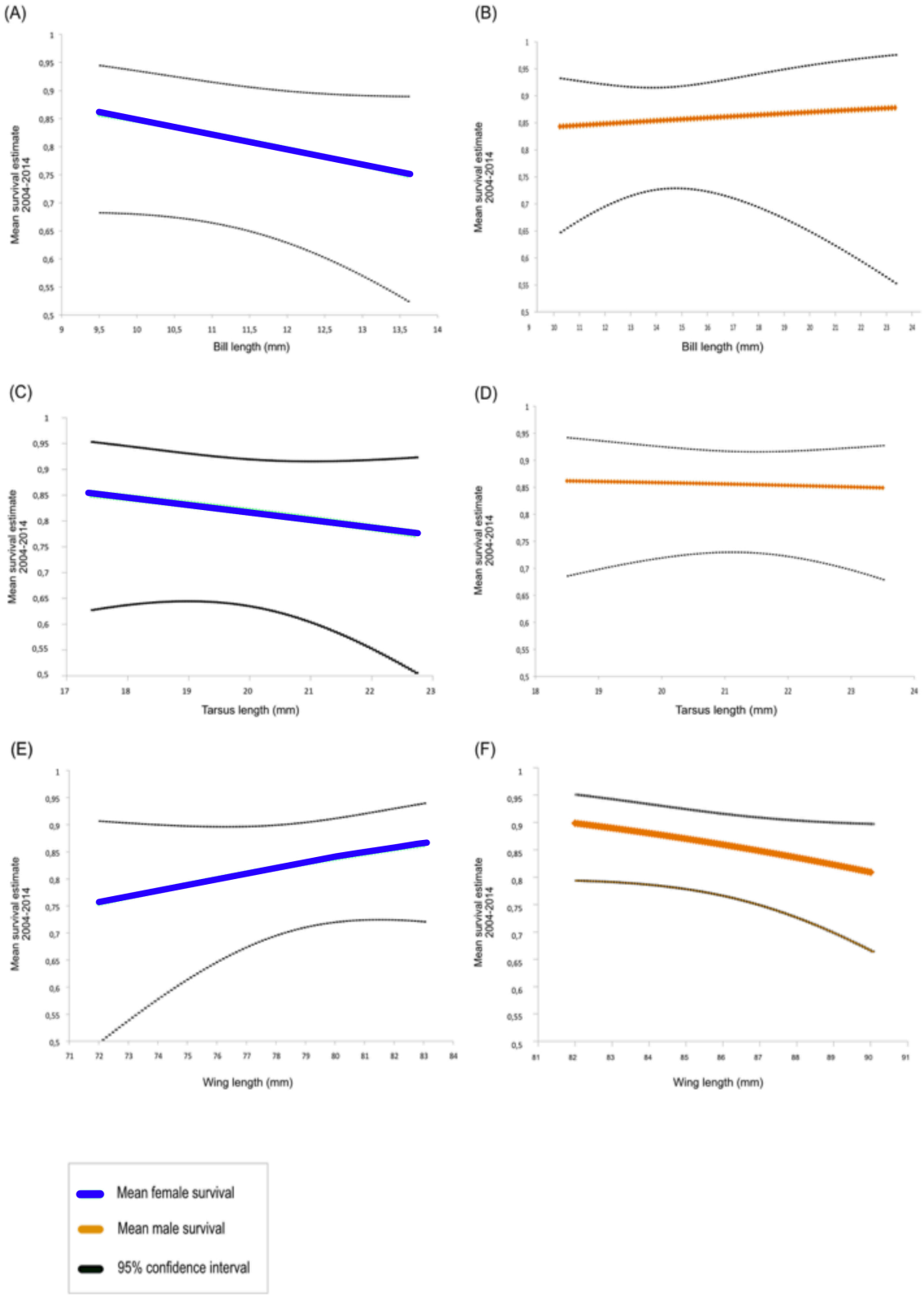


Figure 4.3 Mean survival estimates over the 2004-2014 time period in relation to bill, tarsus or wing length, for females (blue) and males (orange) separately. Estimates are based on the models Phi(t+b) p(.), Phi(t+tar) p(.) and Phi(t+w) p(.). In the case of graphs D and E, these models are not significantly supported by the data.

In the size ratio analysis, the model Phi(t) p(.) ranks the highest. None of the models containing a size ratio variable was supported (Table 4.7). Hence we can confidently reject the hypothesis that size ratio within a mated pair influences survival. Adding up the AIC weights of all the models in which each variable is found, I obtained, for survival, $\Sigma\ AIC_{time} = 1.00$, $\Sigma\ AIC_{bill\ ratio} = 0.20$, $\Sigma\ AIC_{tarsus\ ratio} = 0.15$, and $\Sigma\ AIC_{wing\ ratio} = 0.17$.

Table 4.7 All models tested in the size ratio analysis.

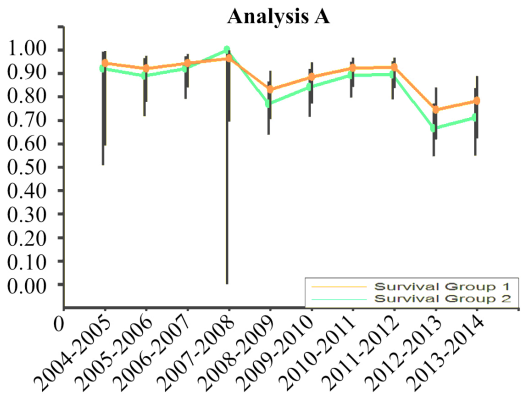
Model	ΔAIC	AIC weight	No. of parameters	Deviance
Phi(t) p(.)	0.0	0.48	11	391.3
Phi(t+B) p(.)	2.0	0.17	12	391.2
Phi(t+W) p(.)	2.1	0.16	12	391.3
Phi(t+TARS) p(.)	2.3	0.15	12	391.4
Phi(t+B+B*t) p(.)	6.4	0.02	21	375.5
Phi(.) p(.)	13.6	0.00	2	423.6
Phi(t+TARS+TARS*t) p(.)	13.7	0.00	21	382.8
Phi(s*t) p(.)	14.0	0.00	21	383.2
Phi(TARS) p(.)	14.1	0.00	3	422.1
Phi(B) p(.)	15.0	0.00	3	423.0
Phi(W) p(.)	15.4	0.00	3	423.4
Phi(t+W+W*t) p(.)	15.5	0.00	21	384.7
Phi(s+TARS) p(.)	16.1	0.00	4	422.1
Phi(s+B) p(.)	17.0	0.00	4	422.0
Phi(s+t+TARS+s*t+TARS*s) p(.)	17.1	0.00	23	381.7
Phi(s+W+s*W) p(.)	17.3	0.00	5	421.2
Phi(t+s+B+s*t+B*s) p(.)	17.4	0.00	23	382.0
Phi(s+W) p(.)	17.4	0.00	4	423.4
Phi(s+TARS+s*TARS) p(.)	18.1	0.00	5	422.0
Phi(s+B+s*B) p(.)	18.6	0.00	5	422.5
Phi(s+t+W+s*t+W*s) p(.)	18.6	0.00	23	383.2
Phi(s+t+B+s*t+B*t+B*s) p(.)	28.8	0.00	32	372.1
Phi(s+t+TARS+s*t+TARS*t+TARS*s) p(.)	29.3	0.00	32	372.6
Phi(s*t) p(s*t)	31.9	0.00	40	355.0
Phi(s+t+W+s*t+W*t+W*s) p(.)	35.7	0.00	32	379.0

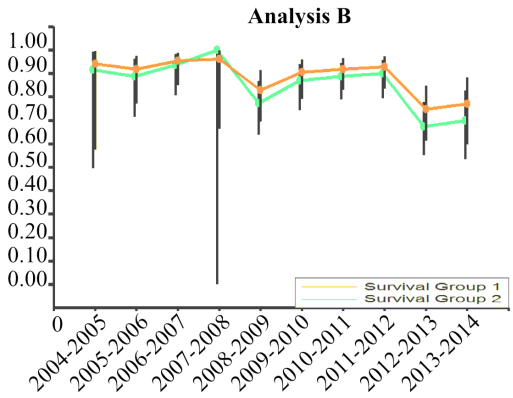
Turning to male genotype analyses, in genotype analysis A, model Phi(t + genotype) p(.) was the most supported, followed closely by model Phi(t) p(.), with a ΔAIC below 2. In analysis B, model Phi(t) p(.) was the most supported, followed closely by model Phi(t + genotype) p(.), with a ΔAIC below 2. In analysis C, the model Phi(t) p(.) was the most supported, with no other model with a ΔAIC below 2. In analysis D, model Phi(t) p(.) was the most supported, followed by model Phi(t + genotype) p(.), with a ΔAIC below 2. In analysis E, model Phi(t + genotype) p(.) was the most supported, followed closely by model Phi(t * genotype) p(.), with a ΔAIC below 2. In analysis F, model Phi(t) p(.) was the most supported, followed by models Phi(t + genotype) p(.) and Phi(genotype2) p(.), both of which have ΔAICs below 2. None of the other models find support in the data (Table 4.8).

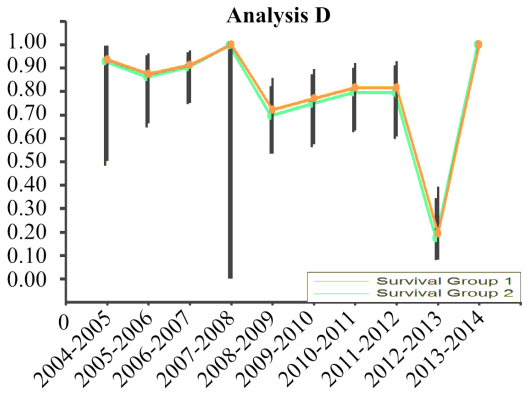
Turning to female genotype, in analysis G, model Phi(.) p(.) was the most supported, followed by models Phi(genotype) p(.) and Phi(t) p(.), both of which generated a ΔAIC below 2. In analyses H and I, models Phi(.) p(.) and Phi(genotype) p(.) were the most supported, the latter model with a ΔAIC below 2 (Table 4.8).

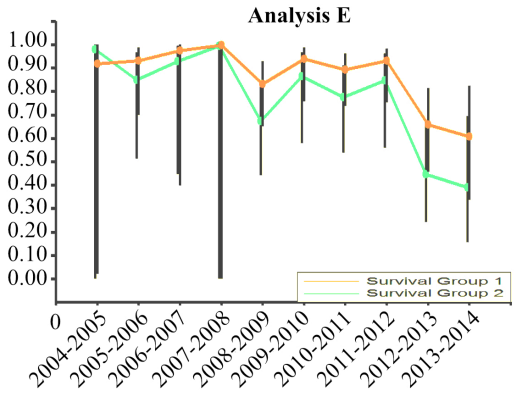
Adding up the AIC weights of all models in which the "genotype" variable is found, I obtained, for survival: $\Sigma\ AIC_{time} = 1.00$ and $\Sigma\ AIC_{genotype}$ values for each analysis can be found in Table 4.8. Two analyses have a $\Sigma\ AIC_{genotype}$ value above 0.50: A and E. The direction of the trends in each analysis is shown in Figure 4.4.

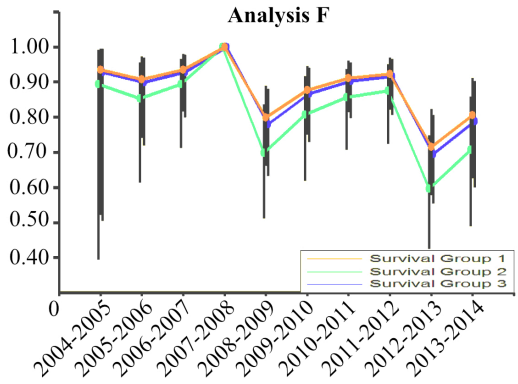
Figure 4.4 (recto) Survival estimates for the groups compared in analyses A-I, excluding analysis C, for which genotype had no effect on survival. The estimates are based on model Phi(genotype+t) p(.) when that model has a ΔAIC below 2 (A-G), and otherwise on model Phi(genotype) p(.) (H-I). The 95% confidence intervals are represented by the black error bars. As explained in Chapter 2, these can sometimes be implausibly large when survival approaches 1, due to MARK's difficulty in calculating such high survival estimates.

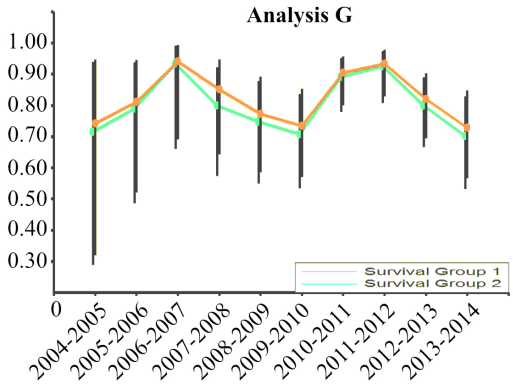

Analysis A
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
2004-2005
2005-2006
2006-2007
2007-2008
2008-2009
2009-2010
2010-2011
2011-2012
2012-2013
2013-2014
Survival Group 1
Survival Group 2

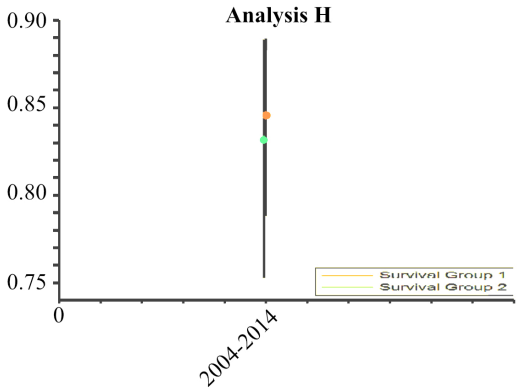

Analysis B
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
2004-2005
2005-2006
2006-2007
2007-2008
2008-2009
2009-2010
2010-2011
2011-2012
2012-2013
2013-2014
Survival Group 1
Survival Group 2


Analysis D
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
2004-2005
2005-2006
2006-2007
2007-2008
2008-2009
2009-2010
2010-2011
2011-2012
2012-2013
2013-2014
Survival Group 1
Survival Group 2


Analysis E
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
2004-2005
2005-2006
2006-2007
2007-2008
2008-2009
2009-2010
2010-2011
2011-2012
2012-2013
2013-2014
Survival Group 1
Survival Group 2


Analysis F
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0
2004-2005
2005-2006
2006-2007
2007-2008
2008-2009
2009-2010
2010-2011
2011-2012
2012-2013
2013-2014
Survival Group 1
Survival Group 2
Survival Group 3


Analysis G
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0
2004-2005
2005-2006
2006-2007
2007-2008
2008-2009
2009-2010
2010-2011
2011-2012
2012-2013
2013-2014
Survival Group 1
Survival Group 2


Analysis H
0.90
0.85
0.80
0.75
0
2004-2014
Survival Group 1
Survival Group 2

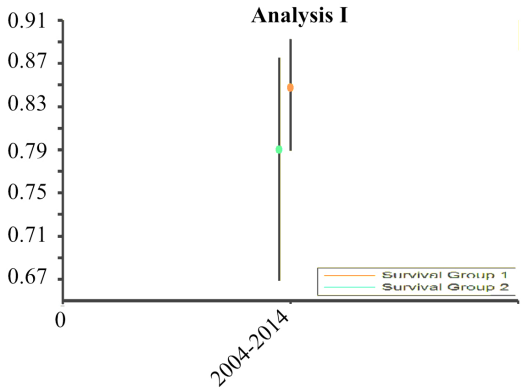

Analysis I
0.91
0.87
0.83
0.79
0.75
0.71
0.67
0
2004-2014
Survival Group 1
Survival Group 2

Table 4.8 All models tested in the genotype analysis with AIC weights ≥ 0.00001, as well as the fully independent model, are included in this table. The notation "genotype" refers to the genotype of all three groups, "genotype1" to the genotype of Group 1, "genotype2" to the genotype of Group 2 and "genotype3" to the genotype of Group 3.

Analysis	Model	Δ AIC	AIC weight	No. of parameters	Σ AIC (genotype)	Deviance
A	Phi(t+genotype) p(.)	0.0	0.54	12	0.54	232.1
	Phi(t) p(.)	0.3	0.46	11		234.6
	Phi(t*genotype) p(.)	14.1	0.00	21		227.3
	Phi(.) p(.)	18.7	0.00	2		271.3
	Phi(genotype) p(.)	19.5	0.00	3		270.1
	Phi(t*genotype) p(t*genotype)	39.9	0.00	40		211.0
B	Phi(t) p(.)	0.0	0.52	11	0.48	234.4
	Phi(t+genotype) p(.)	0.2	0.46	12		232.6
	Phi(t*genotype) p(.)	6.8	0.02	21		219.8
	Phi(.) p(.)	18.2	0.00	2		271.2
	Phi(genotype) p(.)	19.4	0.00	3		270.3
	Phi(t*genotype) p(t*genotype)	30.9	0.00	40		200.4
C	Phi(t) p(.)	0.0	0.74	11	0.26	176.3
	Phi(t+genotype) p(.)	2.1	0.26	12		176.3
	Phi(.) p(.)	13.4	0.00	2		208.2
	Phi(genotype) p(.)	15.4	0.00	3		208.2
	Phi(t*genotype) p(.)	15.8	0.00	21		170.6
	Phi(t*genotype) p(t*genotype)	42.3	0.00	40		153.6
D	Phi(t) p(.)	0.0	0.72	11	0.28	166.6
	Phi(t+genotype) p(.)	1.9	0.27	12		166.4
	Phi(t*genotype) p(t*genotype)	10.9	0.00	40		112.5
	Phi(t*genotype) p(.)	16.2	0.00	21		161.3
	Phi(.) p(.)	54.5	0.00	2		239.6
	Phi(genotype) p(.)	56.5	0.00	3		239.6
E	Phi(t+genotype) p(.)	0.0	0.51	12	0.95	156.9
	Phi(t*genotype) p(.)	0.3	0.44	21		137.8
	Phi(t) p(.)	4.5	0.05	11		163.5
	Phi(genotype) p(.)	12.4	0.00	3		187.9
	Phi(.) p(.)	15.1	0.00	2		192.6
	Phi(t*genotype) p(t*genotype)	27.0	0.00	40		121.0
F	Phi(t) p(.)	0.0	0.44	11	genotype : 0.20	283.7
	Phi(t+genotype) p(.)	1.6	0.19	13	genotype1: 0.06	281.1
	Phi(t+genotype2) p(.)	1.8	0.18	13	genotype2: 0.19	281.3

	Phi(t+genotype1) p(.)	3.2	0.06	13	genotype3: 0.03	282.7
	Phi(t+genotype3) p(.)	4.7	0.03	13		283.6
	Phi(t*genotype) p(.)	5.3	0.00	21		267.8
	Phi(t) p(t)	10.4	0.00	20		275.1
	Phi(genotype2) p(.)	13.5	0.00	3		313.5
	Phi(.) p(.)	15.0	0.00	2		316.5
	Phi(t*genotype) p(t*genotype)	15.4	0.00	21		277.6
G	Phi(.) p(.)	0.0	0.47	2		206.4
	Phi(genotype) p(.)	1.4	0.23	3	0.32	205.8
	Phi(t) p(.)	1.6	0.21	11		189.5
	Phi(t+genotype) p(.)	3.3	0.09	12		189.1
	Phi(t*genotype) p(.)	9.2	0.00	21		175.6
	Phi(t*genotype) p(t*genotype)	32.0	0.00	40		154.9
H	Phi(.) p(.)	0.0	0.70	2		171.5
	Phi(genotype) p(.)	1.9	0.27	3	0.28	171.4
	Phi(t) p(.)	6.7	0.02	11		159.6
	Phi(t+genotype) p(.)	8.7	0.01	12		159.6
	Phi(t*genotype) p(.)	15.7	0.00	21		147.2
	Phi(t*genotype) p(t*genotype)	46.6	0.00	40		134.7
I	Phi(.) p(.)	0.0	0.50	2	0.39	160.0
	Phi(genotype) p(.)	1.0	0.30	3		159.0
	Phi(t) p(.)	2.7	0.13	11		143.8
	Phi(t+genotype) p(.)	3.8	0.07	12		142.7
	Phi(t*genotype) p(.)	14.8	0.00	21		133.5
	Phi(t*genotype) p(t*genotype)	44.5	0.00	40		115.8

Discussion

Because Chapter 3 found very high support for adult survival fluctuating over time, I conducted a MARK analysis to estimate which factors, both year-dependent and individual-dependent, might influence the survival rate of adult Raso larks.

Year-dependent variables

The analyses investigating possible links between environmental, year-dependent factors and survival do not provide support for the hypotheses that population size and population mean clutch size impact adult survival in the Raso lark. Increased population size does not lead to reduced survival in this species, suggesting that intra-specific competition has limited impact on survival. Furthermore, there was

no evidence of an association between the population level clutch size and survival. The analyses do, however, show an effect of yearly rainfall on survival, which makes it the most likely explanation for the inter-annual fluctuations in survival found in Chapter 3. More specifically, the more rainfall in a given year, the lower adult survival in that same year, with a difference of 13% between the wettest and the driest year. This could for example be explained by detrimental changes in feeding conditions for this desert-adapted lark when rainfall increases. A more likely - and somewhat opposite - explanation is that increased rainfall in fact *improves* conditions for the lark, but that this then also results in increased reproductive effort. This might reduce individuals' investment in body maintenance, which in turn reduces their survival. Indeed, in the Raso lark, increases in population size are correlated with higher rainfall. The fact that the analyses in this chapter found no effect between population level clutch size and survival could be due to the very limited and patchy breeding data available.

Body size

In addition to the impact of sex reported in Chapter 3, the other individual-dependent variables were also more strongly correlated with survival than the year-dependent ones. While there did not seem to be any general body size trend associated with survival, some of the morphological measurements taken separately did show such an association. Females with longer tarsi had lower survival rates than females with shorter tarsi, with an estimated 8% difference between the largest and the smallest females. Male survival, on the other hand, was not correlated with tarsus length. It was, however, negatively correlated with wing length: the largest males had an estimated survival rate 9% lower than the smallest males. No such correlation between wing length and survival was found in females. Combined, these tarsus and wing results possibly indicate that, in both sexes, smaller individuals have increased survival, but that, with the available sample sizes, the effect of body size on survival is not strong enough to be systematically detectable for each morphological parameters for both sexes. These differences in body size, especially tarsus length, could be linked to differences in habitat use through foraging modes (Schluter and Smith 1986).

The third morphological measurement, bill size, seems to be a different case, with opposite trends for males and females: while bill size was positively correlated with survival in males, it was negatively correlated in females. Apparently, the effects of bill length on survival work to favour the bill size sexual dimorphism observed in the

species. It is not clear if this is due to inter-sexual feeding competition avoidance. I did not find support for the hypothesis that individuals with a larger size ratio with their mate have increased survival. However, inter-sexual feeding competition avoidance could still be in play, since in very arid years the larks tend to feed in large flocks of both sexes, not in pairs.

Ase18 genotype

In this MARK study, I also investigated the relationship between survival and Ase18 genotype. Since this locus was the only one of the 21 surveyed by Brooke et al. (2010) to have more than two alleles, it might be located next to a gene under selection for variation (Brooke et al. 2010). The most conclusive results were those from analysis E (Σ AIC_{gene} = 0.95) and, to a lesser degree, analysis A (Σ AIC_{gene} = 0.54). Analysis A compared all homozygous males with all heterozygous males. Analysis E compared males homozygous for allele 203 with males heterozygous for allele 203. In both cases, the homozygous individuals seem to have higher survival. The hypothesis that the 203/203 genotype is more advantageous for survival than the 203/- genotype is strongly supported. It is possible that this is also what drives the effect in analysis A, since 203 is the most common allele. This result does not support the heterozygote advantage theory.

To attempt to explain the pattern and gain an insight into the biological function of this locus, I blasted (Altshul et al. 1990) the 189bp Ase18 sequence of the Seychelles warbler *Acrocephalus sechellensis*, for which the Ase18 microsatellite primers were initially developed (Richardson et al. 2000), against the fully sequenced and well-annotated zebra finch *Taeniopygia guttata* genome (Warren et al. 2010). This sequence was 96% similar to a zebra finch sequence at a locus on chromosome 3. According to Ensembl (Flicek et al. 2012), it corresponds to the SYN14 gene, which codes for synaptotagmin XIV, a protein that mediates membrane trafficking in synaptic transmission. In humans, mutations in this gene cause autosomal recessive spinocerebellar ataxia, and a translocation of this gene has been linked to neurodevelopmental abnormalities (Doi et al. 2011). It is possible that this locus also affects the neurodevelopment of the Raso lark; the precise pathway from genotype to phenotype is unknown. Alternatively, a linked locus could be at play.

The additional genotype analyses suggest that allele 203 is the most beneficial for the survival of both males and females, and that allele 224 is the most detrimental,

constantly over time. While this explains why allele 203 is the most common, and might even explain the homozygote advantage, it makes one wonder why all other alleles are still present in the population, especially allele 224, which is the second most common after allele 203 (except allele 205, which all females, but no males, have). This question could be addressed with further research, for example by genotyping more individuals at locus Ase18. Enough blood samples are available to double the sample size of this analysis, which could increase the power of the MARK models.

Age

Finally, the last individual-dependent variable that I analyzed with respect to survival was age (as approximated by "time since marking"). Despite a 10-year long study, age could not be shown to impact survival, and even the first cohort did not show signs of decreased survival over time. This was initially surprising in the context of passerine life history: while for a long time it was thought that senescence could only be observed in captivity due to the many other causes of mortality in the wild, this idea has been refuted in recent years, and a considerable body of literature documenting actuarial senescence in the wild has accumulated (Holmes et al. 2001; Catry et al. 2006). However, many studies, including perhaps ours, are simply too short to cover the onset of senescence in long-lived birds (Péron et al. 2010). In addition, recent studies have shown that birds can maintain high levels of health until old age, with senescence only impacting survival at the very end of the bird's lifespan (Ricklefs 2000; Catry et al. 2006). This phenomenon is heightened in species with slower life histories: the slower the life history, the later the onset of senescence and the slower its pace (Jones et al. 2008; Péron et al. 2010). These explanations fit with the observations of high longevity in the Raso lark and with the picture of a species with a slow life history that is emerging from this study.

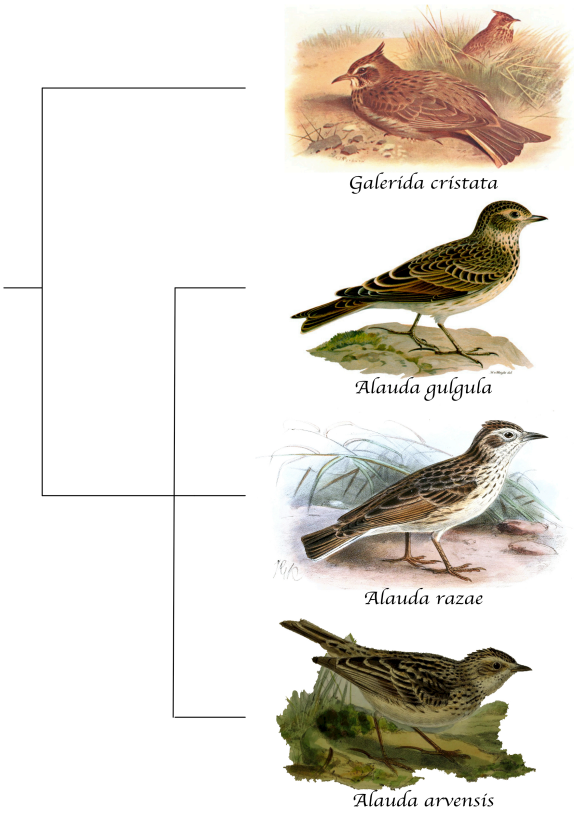
Conclusion

Apart from the somewhat negative effect of increased rainfall in certain years, survival of the Raso lark does not seem to be heavily environmentally driven, and individual covariates such as sex, morphology and genotype seem to be the other key factors, adding to the picture of a species with a high survival rate and heavy investment in maintenance drawn in Chapter 3. The species is one with a long lifespan, high

survival rates (even in the wettest years) and intermittent breeding, potentially only breeding when doing so does not greatly jeopardize survival. This is in line with the high survival rates found in other island passerines, and with current theory on island bird life history. Island birds are thought to be more K-selected than their continental counterparts, with improved adult survival, higher mate fidelity and reduced fecundity (Covas 2012). More resources are shifted towards self-maintenance as a means of increasing survival in order to maximize lifetime reproductive success (Covas 2012). It has been argued that island species are able to do this because of the generally more stable, milder climates on islands and the lower prevalence of parasites and predators (Covas 2012). Both of these arguments are applicable to Raso Island, which has a relatively stable - albeit very dry - environment, very few predators for the lark, and, in all likelihood given its aridity, few lark parasites (Horrocks et al. 2012).

In conclusion, the research in Chapters 3 and 4 sheds light on the Raso lark's life history. Survival significantly varied as a function of year, rainfall, sex, body measurements and genotype. However, this variation does not explain the observed population fluctuations, notably the order of magnitude population increase between 2004 and 2011. Since this increase was not due to immigration, we can conclude that variation in reproductive effort is probably the key driver of population change. That variation is partly manifest in inter-annual variation in clutch size (Table 4.2) but very possibly also in variation in the number of reproductive attempts made by individual larks in the course of a year. Unfortunately the short once-a-year fieldwork visits preclude securing any data on the number of breeding attempts made by any individual lark.

Chapter 5: Phylogeography of the *Alauda* clade



Dendogram of the four lark species studied in this chapter, based on results by Alström et al. (2013). Branch spans carry no significance. Drawings from Wikimedia Commons.

Abstract

A phylogenetic study of all larks by Alström et al. (2013) was not able to determine the precise relationship between the members of the *Alauda* clade: the Raso lark, the Oriental lark and the skylark. The goal of this chapter is to resolve this node on the tree. For this purpose, I used RAD sequencing and sampled the skylark across its broad geographical range. Based on my results, the Raso lark and the skylark are sister species, and the Oriental lark branch should be collapsed with the skylark. The Oriental lark appears to be a subpopulation, or maybe a subspecies, of the skylark. I estimated a coalescence time of 6.4 million years between the crested lark and the *Alauda* clade, and a coalescence time of 5 million years between the Raso lark and the other *Alauda* species. These numbers are in line with the split times estimated by Alström et al. (2013) of 8.5 and 5.5 million years ago respectively. Additionally, this chapter also introduces the whole genome of the skylark.

Introduction

The family Alaudidae is a clearly defined and delimited group, based on unique tarsus and syrinx features (Donald 2004; Alström et al. 2013). The syrinx - the voice organ unique to birds situated at the junction of the trachea and bronchi - is characterized in larks by the absence of a bony pessulus and the presence of only five sets of muscles instead of the six to eight sets found in other oscine songbirds. The tarsus of larks has a rounded back edge and is covered in small scales instead of the larger, flatter scales found in other songbirds (Donald 2004).

The designation of genera within this family, however, is much more debated (Harrison 1966; Donald 2004; Alström et al. 2013): their number oscillates between 19 and 23 (Donald 2004). The reason for this is that, traditionally, the designation of these genera has been based mainly on bill structure and plumage, two characteristics that are very much diet- and habitat-dependent and, as such, unreliable for phylogenetic studies. While variations in plumage coloration have strong biological implications for birds in general, it is particularly so for larks, a family that mostly lives in open habitats and for which cryptic plumage is crucial to survival. The colour of larks is extremely adaptable to local substrate and vegetation colour (Donald 2004; Alström et al. 2013). As a result, the Alaudidae family comprises several monotypic genera (e.g. *Lullula*; Figure 5.1) and enigmatic genera that have challenged taxonomists (Alström et al. 2013). For example, the genera *Galerida, Alauda* and *Lullula* appear to be very close, and some authors have suggested grouping them into a single genus (Donald 2004).

Even at the species level, lark taxonomy based on morphology is difficult. Some authors are of the opinion that recent molecular advances will lead to the discovery of new species (Donald 2004; Alström et al. 2013). The classification of certain lark species, such as the Raso lark (see Chapter 1), has been debated for decades. Differentiating between the members of the *Alauda* clade based on their very similar morphology is not straightforward, especially given these species' high levels of within-species variation (Figure 5.2, Table 5.1). Based on the Alström et al. (2013) study, the white-winged lark *Melanocorypha leucoptera* was recently added to the *Alauda* clade and renamed *Alauda leucoptera* (Figure 5.1). In this thesis, however, the term "*Alauda*" only refers to the skylark, Oriental lark and Raso lark, for the sake of consistency with older sources.

Indeed, intraspecific lark phylogenetics based on morphology is problematic. While the Raso lark, thanks to its tiny range and population, does not cause such

troubles to the ornithologist, its two closest relatives, the skylark and the Oriental lark (Figure 5.1), certainly do. Their local adaptations, coupled to their large ranges (Table 5.1), result in extreme within-species morphological variation.

The skylark's different forms differ mainly in bill size, body size and ground colour (Figure 5.2). Donald (2004) lists 12 generally recognized subspecies. There are no geographic trends in bill size and shape, except that North West African individuals have proportionally the heaviest and longest bills (Campbell et al. 1997). Body size appears to increase from western Europe to western Siberia, where it reaches a maximum, and then decreases again towards the Pacific. The variations in plumage colour are also part of a gradient: neighbouring forms often grade into each other and differences in colour between adjacent forms are very slight - smaller than between individuals within the form. As a result, although these different forms of the skylark are generally treated as subspecies, they are almost impossible to distinguish in the field, and their subspecific status may not be warranted (Donald 2004).

The Oriental lark also displays substantial morphological variation, with at least ten forms being regularly cited, and many more proposed in the literature. For example, the largest and palest Oriental larks tend to be found amongst the westernmost populations, *Alauda gulgula inconspicua* (Donald 2004). However, similarly to the skylark, counting and delimiting Oriental lark subspecies geographically based on morphology seems to be at an impasse (Meinertzhagen 1951; Donald 2004; Alström et al. 2013).

Despite the limitations of morphology-based phylogenetics in larks, only two molecular phylogenies have been published so far. The first study looked at a small number of mostly African species - but not the Raso lark - using mitochondrial DNA (Tieleman et al. 2003). The second study was much larger in scope and comprised 81 of the 97 lark species (Alström et al. 2013). It used two mitochondrial and three nuclear loci for the skylark, one mitochondrial and three nuclear loci for the Oriental lark, and two mitochondrial loci for the Raso lark. It found a high level of discrepancy between classifications based on morphology and on DNA, to an extent found in very few other bird groups. The authors' results show that there are remarkable morphological similarities between distantly related species and, conversely, strong divergence between some sister species, most notably for traits related to feeding (size and shape of bill) and plumage. Additionally, of specific interest to this thesis, Alström et al. (2013) refuted the suggestion that *Alauda, Galerida* and *Lullula* might be grouped together into one genus (Donald 2004; Figure 5.1). Furthermore, they confirmed that the Raso lark

belongs in the genus *Alauda* (Figure 5.1). However, they were not able to determine the precise phylogenetic relationship between the Raso lark, the Oriental lark and the skylark; this node on the Alaudidae tree remained unresolved (Figure 5.1). Their data non-conclusively suggest that the skylark and the Oriental lark might be sister species (Alström et al. 2013).

The goal of this chapter is to test this hypothesis and, if possible, resolve this node on the Alaudidae tree. To this purpose, I used RAD sequencing, which provides a larger number of loci and better reflects variation across the whole genome (Brito and Edwards 2009; Edwards and Bensch 2008; Dierickx et al. 2015), sampled the skylark more broadly geographically to account for its intraspecific variation (Figure 5.3), and included an outgroup of the *Alauda* clade, the crested lark *Galerida cristata* (Figure 5.1). Resolving this node is important for Raso lark research, since it will establish which species is its closest relative and hence set the correct points of reference for, amongst others, the study of both its conserved and derived traits. More generally, resolving this node means setting a robust phylogenetic framework for future genetic, phenotypic or ecological studies of the Alaudidae, including potentially the quantification of the evolutionary rates for specific traits in the *Alauda* larks.

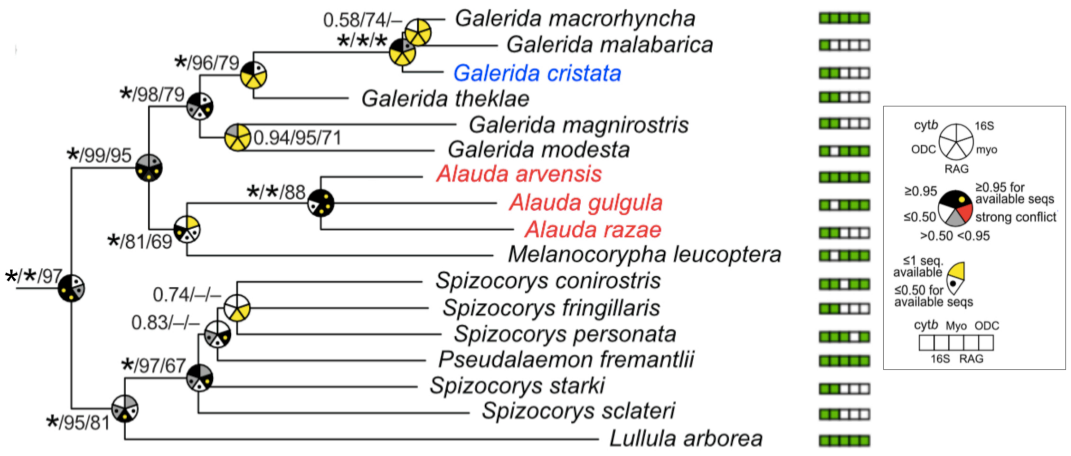
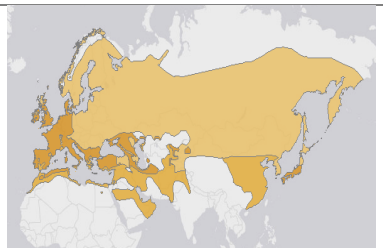
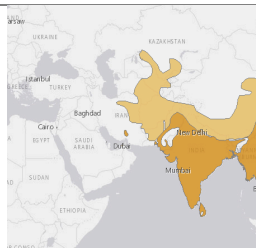


Figure 5.1 Majority rule (50%) consensus tree based on concatenated nuclear ODC, myoglobin and RAG1 + 2 and mitochondrial cytochrome b (cytb) and 16S sequences, inferred by Bayesian inference, analysed in five partitions (one per locus; all mixed + C + I). Pie charts indicate posterior probabilities in single-locus analyses (see legend on right). Support values are indicated at the nodes, in the order posterior probability/maximum likelihood bootstrap/parsimony bootstrap; an asterisk represents support 1.0/100%. Coloured boxes to the right indicate sequences available for each species (see legend on right). Species in red are the focus of this chapter; the species in blue is the outgroup (adapted from Alström et al. 2013).

Figure 5.2 Morphological variation in the *Alauda* genus (left plate, top to bottom): skylark, Japanese skylark (previously thought a separate species), Oriental lark and Raso lark; morphological variation in the *Alauda arvensis* species (right plate, clockwise from top left): adult *arvensis*, juvenile *arvensis*, adult *cantarella*, adult *sierrae*, adult *pekinensis*, egg variation, adult *harterti*, adult *dulcivox* (copied from Donald 2004).

Table 5.1 Morphological and behavioural comparison between the Raso lark, skylark, Oriental lark and crested lark, showing their great interspecific similarity and intraspecific variation (Brooks et al. 1988; Flegg and Hosking 2002; Donald 2004; Donald and Brooke 2006; IUCN 2016). Darker orange shading on the maps indicates where the species occurs throughout the year; paler orange indicates areas occupied only during the breeding season.

	Raso lark *Alauda razae*	Skylark *Alauda arvensis*	Oriental lark *Alauda gulgu*
Range (dark orange: resident; pale orange: breeding)			
Within-species variation	None	Numerous subspecies	Numerous subspecies
Body size	30% smaller than the skylark	Increases from western Europe to a maximum in western Siberia, and then declines towards the Pacific	15% smaller than the skylar and tail
Plumage colour	Greyer, lacking rufous tones. The chicks do, however, have these tones	Streaked brown; considerable variation throughout the range	Streaked brown; considerab throughout the range
Song	Well developed and prolonged; possibly less melodious and more monotonous than the skylark	Well developed and prolonged: a high, liquid “chirrup”	Well developed and prolong buzzing “baz baz,” but also skylark calls
Flight display	Reach great heights; gentle wing beats; vertical ascent; parachute to the ground	Reach great heights; open their tail; spiral ascent; parachute to the ground	Reach great heights; keep th descend in steps
Morphology	Less pointed wing than the skylark	Encapsulates the form of the family for most observers: fairly stout and long body, quite long legs, wings and tail	More delicately proportione skylark; proportionally long upright stance
Bill	Heavier, longer than the skylark; sexually dimorphic: larger in males than in females	Unspecialized, relatively strong conical bill	Unspecialized conical bill; s and thinner than the skylark
Crest	Erectile crest	Erectile crest	Erectile crest
Diet	Omnivore; drink infrequently or not at all	Omnivore; drink infrequently or not at all	Omnivore; drink infrequentl
Behaviour	Resident; flock outside of breeding season	Migratory; flock outside of breeding season	Migratory; flock outside of
Habitat	Arid	Temperate to semi-arid; closely associated with farmland; unspecialized open habitats	Temperate to semi-arid; clos with farmland; unspecialize

Methods

Field and laboratory work

Blood samples from 33 Raso larks were collected non-destructively and preserved following the methods outlined in Chapter 2. From colleagues, I also obtained blood and tissue samples of skylarks, Oriental larks and crested larks from six different Eurasian populations: 15 skylarks from the Netherlands, 10 skylarks from Western Russia (WR), 11 skylarks from Eastern Russia and Mongolia (ERM), eight skylarks from China, 11 Oriental larks from Taiwan and nine crested larks from Saudi Arabia (Figure 5.3). None of the sampled birds was likely to be a migrant based on the sampling date (Table 5.S1) and/or the migration pattern of the species (BirdLife International 2016; IUCN 2016). More details on the sampling methods, including exact location, tissue type and collection date can be found in Supplementary Materials (Table 5.S1). These samples were preserved in ethanol or TE buffer and stored at -80°C.

DNA was extracted from the blood and tissue samples using the Qiagen DNeasy kit (Qiagen, Venlo, the Netherlands), following the manufacturer's protocol for tissue samples, except for an increased digestion time (overnight). DNA concentrations were measured with fluometric quantification (Qubit 2.0 HS DNA assay; Invitrogen, Life Technologies, Carlsbad, CA, USA). Single digest RAD sequencing libraries for each individual sample (except individual 0) were prepared according to a protocol developed by the Butterfly Genetics Group at the University of Cambridge (Merrill 2014) using the enzyme *PstI-HF*. Each individual was assigned an 8-base pair (bp) inline barcode, and equimolar concentrations of 16 uniquely barcoded individuals were pooled and double-indexed by 16 cycles of high-fidelity PCR using Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific, Waltham, MA, USA) with Illumina barcodes, resulting in a unique combination of inline and Illumina barcodes for each individual. The PCR products were pooled in equimolar quantities and sequenced in four lanes of two rapid runs of an Illumina HiSeq 1500 at the Gurdon Institute facility at the University of Cambridge, producing 100-bp single-end reads.

In collaboration with Simon Yung Wa Sin, a draft reference genome was obtained through the whole-genome sequencing of a male skylark sample, individual 0, collected in Mongolia (Harvard MCZ 349891; Table 5.S1). Two libraries were prepared: a 220bp insert size fragment library using a PrepX ILM 321 DNA kit for Apollo 324 and following the manufacturer's protocol (IntegenX, Pleasanton, CA,

USA), and a 3KB mate pair library using an Illumina Nextera Mate Pair Sample Preparation kit and following the manufacturer's protocol (Illumina, San Diego, CA, USA). These two libraries were sequenced with two high output runs on an Illumina HiSeq 2500 at the Bauer Core facility at Harvard University, producing 125bp paired-end reads.

Figure 5.3 Sampling locations of the seven lark populations. Sample sizes of sequenced individuals for each population are given below each point in black, and sample sizes for each population after removing poorly sequenced and related individuals are given in red. See main text for removal criteria.

Skylark whole genome assembly

The skylark genome for individual 0 (Table 5.S1) was assembled following the methods outlined in Gnerre et al. (2011) and the Illumina 2500 platform. Trimmomatic 0.32 (Bolger, Lohse and Usadel 2014) was used to trim the Illumina adaptors from the paired-end reads, FastQC (Andrews 2010) to check read quality and Allpaths-LG (Gnerre et al. 2011) to assemble the genome. Genome quality and summary statistics were also calculated with Allpaths-LG. The assembled genome was then prepared for alignment with the RAD sequences with the *genome_indexing* and *genome_inspecting* modules in the program Bowtie 2 (Langmead and Salzberg 2012).

Sequence processing and alignment

I used the program *process_radtags* in Stacks 1.35 (Catchen et al. 2013) without any quality filters to sort sequence reads by barcode. I then used Trimmomatic to crop all reads to a maximum length of 95bp and remove all reads that were shorter than 95bp, thereby ensuring that all reads were of equal length, and to trim the 6bp corresponding to the restriction sites. *Process_radtags* was then used one more time to process the sequences and filter them for quality. A *de novo* assembly was performed with *denovo_map.pl* in Stacks using parameters –m 5 –M 4 –n 3 –p 2. All individuals were also aligned to the skylark genome using Bowtie 2. Reads with multiple significant matches to the reference genome were removed, ensuring that each read corresponded to only one alignment. I then ran the *ref_map* module in Stacks.

At this stage, I eliminated all individuals with more than 60 percent of unpopulated sites across the total dataset (three Raso larks, three crested larks and one skylark from China). I also ran the program KING (Manichaikul et al. 2010) with option --*homo* to calculate relatedness between individuals, and I randomly removed one individual from each potential first degree-related pair (three Raso larks, one crested lark, three skylarks from China, one skylark from the Netherlands and one skylark from Western Russia), since this would bias the analyses (Figure 5.3). Finally, I used *populations* in Stacks to create data matrices (referred to as "datasets," Table 5.2) for subsequent analyses, with an individual minimum locus stack depth of either 10 or 20. To study the effects on genetic estimates of different stack depths, different alignment methods, different SNP filtering protocols and removal of low-frequency alleles, I analyzed nine different datasets (numbered 1 through 9) produced from different alignment and filtering protocols, varying minor allele frequency (MAF) and heterozygosity requirements (Table 5.2). Specifically, I varied the number of SNPs per locus (to avoid SNPs in potential high linkage disequilibrium), removal of SNPs with heterozygosity > 0.75 and a MAF < 0.05 (to avoid potential low-frequency SNP miscalls) and dataset completeness.

Table 5.2 Description of datasets 1-9 and the SNP filtering parameters used to create them.

Dataset number	Stack depth required for individuals at a locus (-m)	SNP filtering	Dataset completeness requirement for locus inclusion	Dataset completeness† (%)	Total number of SNPs	Total number of loci
Alignment method: reference genome						
1	10	no	At least 50% of individuals in at least 2 populations	50	15005	1964
2	20	no	At least 50% of individuals in at least 3 populations	64	4892	629
3	10	1 SNP per locus	At least 50% of individuals in at least 2 populations	42	1900	1900
4	20	1 SNP per locus	At least 50% of individuals in at least 3 populations	64	604	604
5	10	1 SNP per locus, loci required to have heterozygosity < 0.75 and MAF > 0.05	At least 50% of individuals in at least 2 populations	40	1731	1731
6	20	1 SNP per locus, loci required to have heterozygosity < 0.75 and MAF > 0.05	At least 50% of individuals in at least 3 populations	33	341	341
Alignment method: *de novo*						
7	10	no	At least 50% of individuals in at least 2 populations	37	30941	4984
8	10	1 SNP per locus	At least 50% of individuals in at least 2 populations	36	4748	4748
9	10	1 SNP per locus, loci required to have heterozygosity < 0.75 and MAF > 0.05	At least 50% of individuals in at least 2 populations	33	4188	4188

† Average proportion of populated sites across the total dataset for all individuals.

Population differentiation

To assess population differentiation between all the lark populations, I calculated pairwise F_{ST} values among the seven populations with Stacks for all datasets. Similarity between the resulting F_{ST} matrixes for the different datasets was evaluated with a Mantel test run in R 3.2.4. To examine population differentiation in a multivariate framework, I used R to run a discriminant analysis of principal components (DAPC) on datasets 1-4, 7 and 9 using the *adegenet* package (Jombart and Ahmed 2011). For additional insights into the population structure of the different lark populations, I used a Bayesian approach implemented in the program STRUCTURE 2.3.1 (Pritchard et al. 2000) on datasets 3 and 8 containing all seven populations to determine the number of genetic groups, or "clusters" (K) that best fit the data, and assign individuals to cluster(s). Because the higher levels of population structure can hide lower levels of structuring (Evanno, Regnaut and Goudet 2005), I then took a nested approach, first removing the most distant outgroup, the crested lark, and then also the Raso lark, in order to zoom in first on the skylark/Raso lark/Oriental lark phylogenetic node, and then on the skylark/Oriental lark complex. STRUCTURE was run three times for each value of K ranging from 1 to 7 under an assumption of admixture, with 200,000 cycles of burn-in (BURNIN = 200000) and 400,000 Markov chain Monte Carlo samples (NUMREPS = 400000). I combined the replicate result file using CLUMPP 1.1.2 (Jakobsson and Rosenberg 2007), implemented in STRUCTURE HARVESTER (Earl and vonHoldt 2012), and visualized the combined result file with *distruct* 1.1 (Rosenberg 2004). All datasets used for STRUCTURE were datasets with only one SNP per locus, since linked loci can affect STRUCTURE results.

Phylogenetic trees and coalescence time

I built two different types of concatenated phylogenetic trees of the lark populations, one based on a cladistic method (maximum likelihood) and one based on a phenetic method (neighbour-joining), using the crested lark as the outgroup. The maximum likelihood (ML) tree was built using TreeMix 1.13 (Pickrell and Pritchard 2012), which is a program that, if given a set of allele frequencies from a number of populations, will return the maximum likelihood tree for the set of populations and identify populations that are poor fits to the tree model. Then, optionally, it can model migration events involving these populations to improve the fit. The neighbour-joining

(NJ) tree was built with the R package *adegenet* using Prevosti's distance (Prevosti 1975) - a measurement over all loci of the proportion of unshared alleles - and 1000 repetitions for the bootstrapping values. I explored numerous data filtering options for different analyses (STRUCTURE, DAPC and F_{ST}), but because it made little difference to the results, I based the phylogenetic trees on my reference dataset, dataset 1.

In order to put the phylogenetic trees into a timeframe, I calculated the coalescence times between the different species. To do this, in collaboration with Simon Martin, we first calculated the mean weighted absolute divergence (dxy) between each pair of populations using a python script (*popgenWindows.py* available at https://github.com/simonhmartin/genomics_general/). This was performed on dataset 1, but with relaxed dataset completeness requirements for site inclusion: a site needed to be present in at least of 25% of individuals to be included. This was done to maximize the number of sites per window, since this script works on 100 kb windows. Windows are rejected if they have fewer than 100 sites meeting the site inclusion requirements. I then used these values of dxy to calculate the coalescence time between the crested lark and the *Alauda* clade, and between the Raso lark and the other *Alauda* populations, based on the relationship

$$t = \frac{dxy}{2\mu}$$

where t is the average coalescence time between all pairs of haplotypes in two populations, dxy is the absolute divergence between these two populations, and μ is a mutation rate of 1.5×10^{-9} per site per year in birds (Ellegren et al. 2007).

Results

Characterization of the skylark genome

The sequencing of the fragment and the mate pair libraries for the skylark genome produced 154.34 million and 263.95 million reads respectively. The assembly was based on 36,732 contigs and 5,714 scaffolds. The N50 contig size was 71.5 kb (Figure 5.4). The N50 scaffold size was 1,444 kb. The estimated genome size of the skylark was 1.06 billion base pairs. The GC content of the fragment reads was 42.9%. Eleven percent of the genome was estimated to be repetitive.

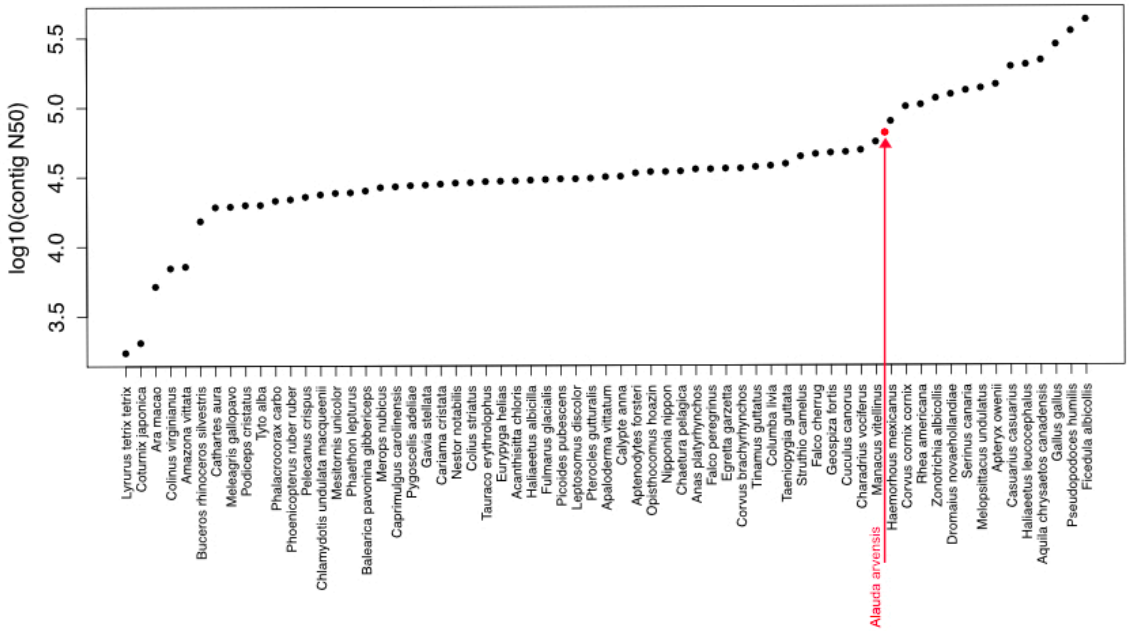


Figure 5.4 The contig N50 (an indicator of quality) of the skylark genome added in red onto a graph produced by Tim Sackton in 2016 (personal communication) to show how the skylark contig N50 compares to the other sequenced avian genomes.

Characterization of RADseq data

Sequencing produced 482,421,923 single-end reads across 96 individuals in the seven lark populations. After *process_radtags* filtering in Stacks, the catalogue contained a raw total of 14,585,276 reads. Dataset completeness, defined as the average proportion of populated sites across the total dataset for all individuals, ranged from 33% to 64% depending on the filtering procedures in *populations* specific to each dataset (Table 5.2). In dataset 1, the reference dataset against which I tested other read-processing procedures, 1,964 loci were retained after filtering in *populations* for a minimum stack depth of 10 and presence in at least 50% of the individuals of two populations (Table 5.2). In this set of loci, I found a total of 15,005 SNPs (variable sites). The resulting data matrix was 50% complete across all SNPs and individuals (Table 5.3).

Table 5.3 Sequencing run information per population (crested lark, Raso, The Netherlands, Western Russia, Eastern Russia and Mongolia, China, Taiwan or all) for all positions (both variant and fixed), based on dataset 1 (in black) and dataset 7 (in blue).

Populations	All
Total number of reads produced by the RADseq run	482421923
Number of loci in Stacks catalogue before filtering in populations	3709023
Number of loci retained in Stacks catalogue after filtering in populations	1964
Number of detected SNPs	15005
Mean stack depth of coverage	52.6

Populations	Crested lark	Raso	The Netherlands	Western Russia	Eastern Russia and Mongolia	China	Taiwan
Average number of individuals per population across loci	4.2 4.0	22.8 20.6	10.6 9.4	7.7 6.9	7.9 7.2	2.9 2.7	9.6 8.4
Total number of polymorphic RAD sites present in all individuals	1786 3181	3008 5451	6780 10966	2498 3888	4283 7939	4037 7193	4009 6720
Number of variant sites	6748 12055	8529 15411	11561 19366	6860 11571	9834 18341	12716 22847	9800 16280
Mean missing data (%)	61.5 70.2	52.8 64.1	41.8 60.3	62.6 72.9	48.2 59.2	37.1 52.4	44.2 62.0

Population differentiation

Pairwise F_{ST} values were highly similar for all datasets, showing robustness to different alignment and SNP filtering methods. Mantel tests between the result matrixes for the different datasets allowed rejection of the null hypothesis that the matrixes were not correlated at the 0.05 level, and found correlation coefficients > 0.9. As such, I only report the results for dataset 1. $F_{ST\text{-}ERM\text{-}WR}$, $F_{ST\text{-}ERM\text{-}Netherlands}$ and $F_{ST\text{-}WR\text{-}Netherlands}$ were the lowest, with values ≤ 0.05 (0.05, 0.04 and 0.03 respectively). The pairwise F_{ST} values that included the outgroup, the crested lark, were the highest amongst all pairwise comparisons (≥ 0.18). The pairwise F_{ST} values that included the Raso lark were the second highest ($0.10 < F_{ST} < 0.14$, excluding $F_{ST\text{-}crested\ lark\text{-}Raso}$), with $F_{ST\text{-}Raso\text{-}}$

$_{Netherlands}$ being the lowest in this category ($F_{ST\text{-}Raso\text{-}Netherlands}$ = 0.10). The China and Taiwan populations had pairwise F_{ST} values with each other and with the skylark populations that ranged between 0.06 and 0.08 (Table 5.4).

Table 5.4 Pairwise F_{ST} values based on dataset 1. Darker colours indicate higher values; corresponding intervals are [0, 0.05], [0.06, 0.10], [0.11, 0.15] and > 0.15.

dataset 1	Raso	China	Taiwan	Western Russia	Eastern Russia and Mongolia	The Netherlands
Crested lark	0.26	0.29	0.24	0.24	0.24	0.18
Raso		0.14	0.13	0.13	0.12	0.10
China			0.08	0.08	0.06	0.06
Taiwan				0.06	0.06	0.06
Western Russia					0.05	0.03
Eastern Russia and Mongolia						0.04

For datasets 1-4, 7 and 9, using the first 10 PC axes and five discriminant axes, the DAPC plot shows that crested lark, Raso lark and Oriental lark individuals cluster as three distinct populations, while skylark individuals from the Netherlands, West Russia, East Russia & Mongolia and China all cluster together (Figure 5.5). Only plots based on datasets 3 and 4 show the Oriental lark clustering with the skylark populations; however, this is probably an artefact of the low total number of loci in these datasets (Table 5.2).

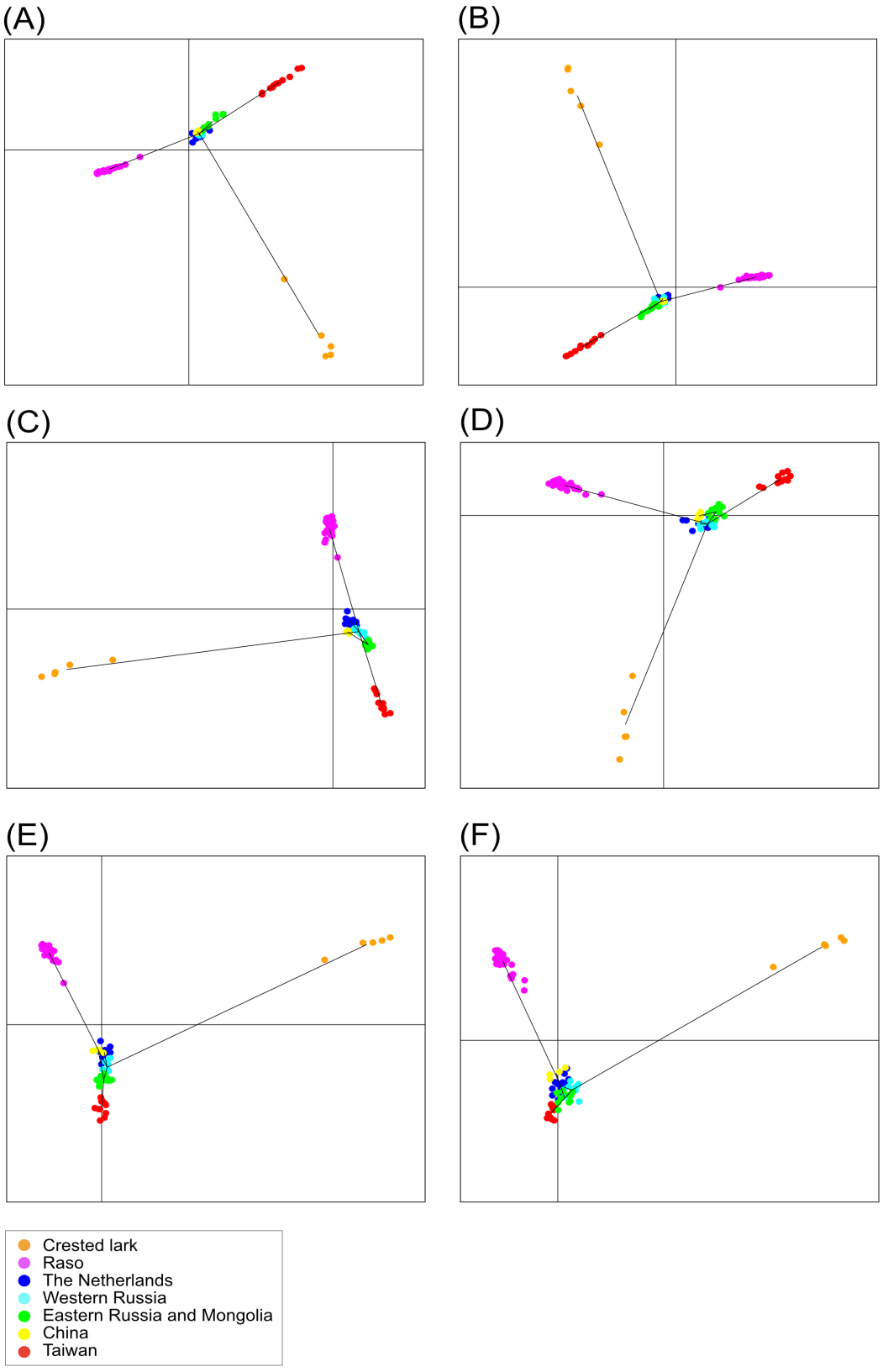


Figure 5.5 DAPC plots based on two *de novo* alignment datasets, dataset 7 (A) and dataset 8 (B), and four reference genome alignment datasets, dataset 1 (C), dataset 3 (D), dataset 2 (E) and dataset 4 (F). The horizontal axis corresponds to the first discriminant, the vertical axis to the second. The black lines linking the different populations indicate to which other population each population is closest.

In the STRUCTURE analysis, the mean log-likelihood across the three runs for each value of K was found to be maximized at K = 4 clusters. The crested lark and the Raso lark always formed distinct clusters. Skylarks from the Netherlands and from Western Russia were very similar. The Oriental larks from Taiwan shared considerable genetic variation with the skylark populations, but were still distinguishable from the skylarks from the Netherlands and Western Russia. The skylarks from China and Eastern Russia and Mongolia had a genetic profile between that of the Taiwanese population and of the most western populations, reflecting their geographic location. These results were even more pronounced when crested larks were excluded from the analysis: the Oriental lark now shared almost no genetic variation with the Dutch and Western Russian skylarks, and the skylarks from China and Eastern Russia and Mongolia were again intermediate (Figure 5.6).

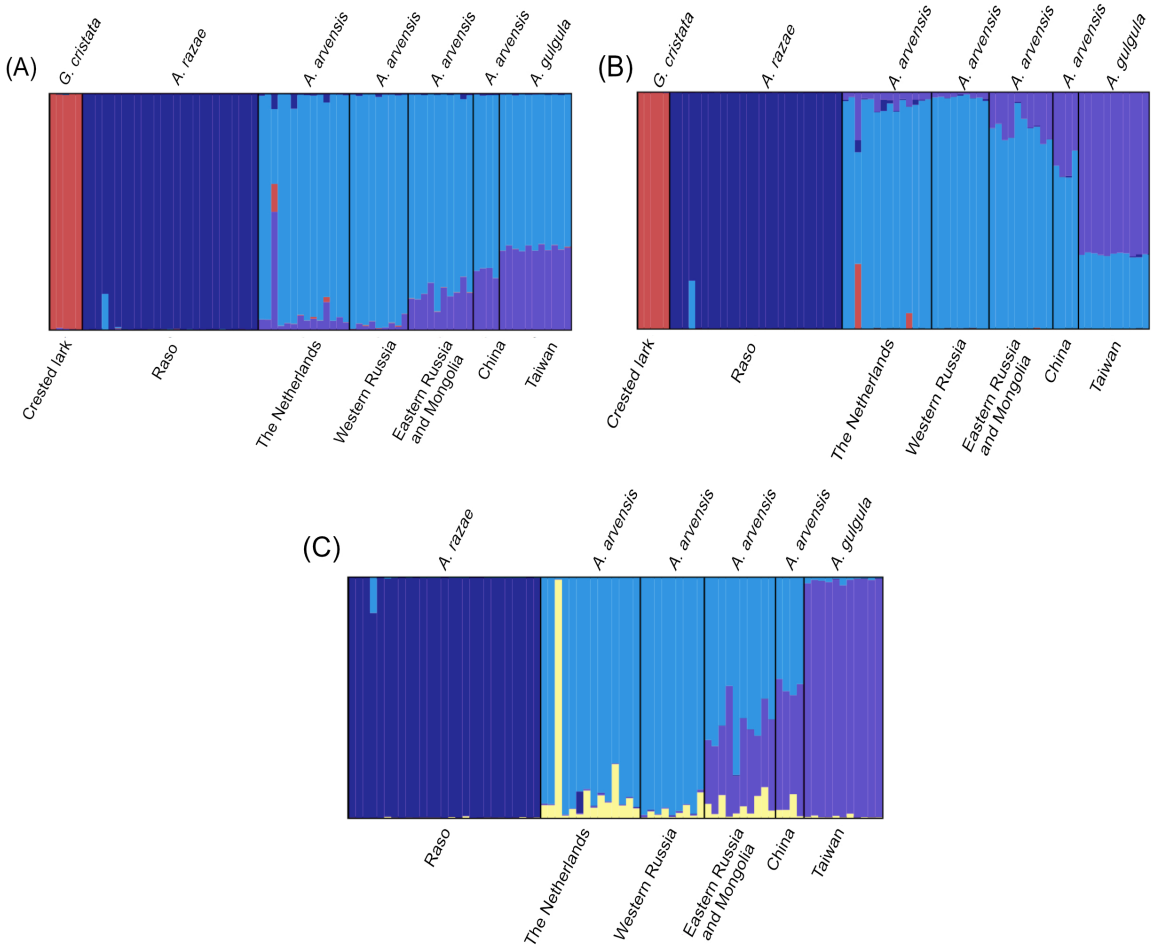


Figure 5.6 STRUCTURE figures based on datasets 3 (A) and 8 (B) including all lark species, and on dataset 3 excluding the crested lark (C). The populations are ordered according to their geographical location, from west to east. Models with K = 4 received the most support for all three datasets. It is unclear what causes one sample in the Dutch population to appear as an outlier, maybe laboratory or sequencing contamination.

Phylogenetic trees and coalescence time

The Raso lark is clearly located on a separate branch on both the ML tree (Figure 5.7) and the NJ tree (Figure 5.8). On the NJ tree, the bootstrap value for the node separating the Raso lark from the Oriental lark and the skylark is 100. The NJ tree places the Taiwan population of Oriental larks within the different skylark populations, with support values for the branches separating the different skylark and Oriental lark individuals well below 50 in most cases (Figure 5.8). The ML tree places all skylark and Oriental lark populations on the same branch. The ML tree's drift parameter also shows that the Raso lark has undergone more genetic drift than the skylark, which is expected given its much smaller population size (Figure 5.7). The residuals for the ML tree

showed adequate fit between the data and the tree model, and did not suggest a need to add admixture events to the model (Figure 5.S1).

The highest values of absolute divergence between pairs of populations was found between the crested lark (the outgroup) and the other populations, with dxy ranging from 0.0187 to 0.0197. This was followed by the divergence between the Raso lark and the other *Alauda* populations, with dxy ranging from 0.0149 to 0.0152. The other *Alauda* populations, including the Oriental lark population in Taiwan, all had lower values of divergence with each other, and the divergence generally corresponded to the geographical distance between the populations: dxy = [0.0123, 0.0146] (Figure 5.9). Because all dxy values involving the crested lark were very similar, I used the mean of these values (0.0193) to calculate a coalescence time between the crested lark and the *Alauda* clade of 6.4 million years. Similarly, I used the mean of the dxy values involving the Raso lark and the other populations in the *Alauda* clade (0.0150) to calculate a coalescence time between the Raso lark and its sister species of 5 million years.

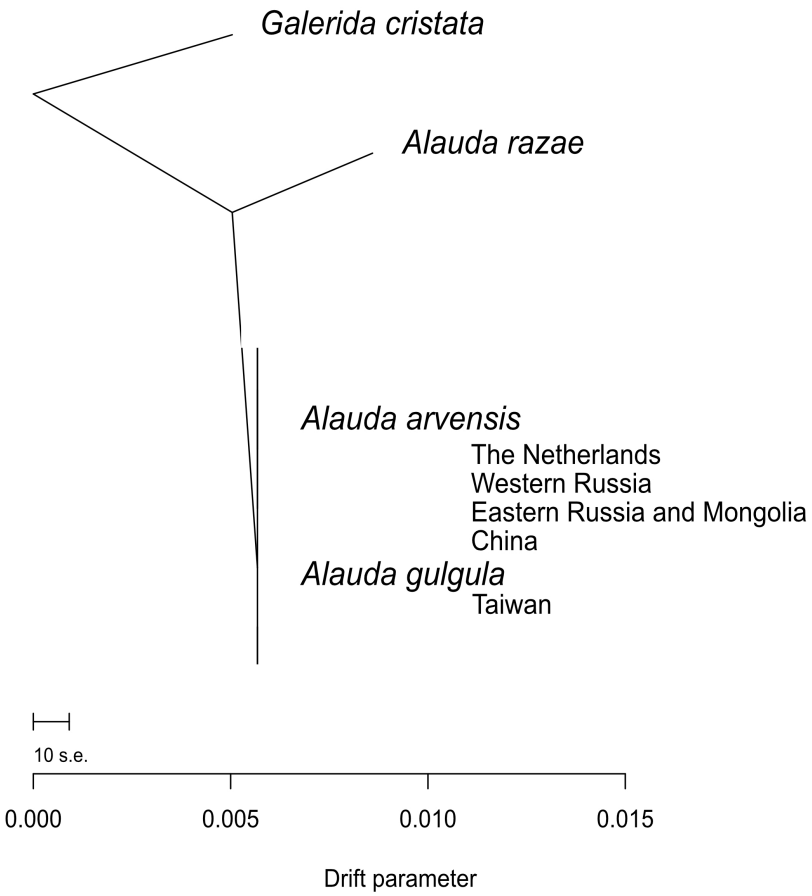


Figure 5.7 Maximum likelihood tree built using TreeMix. The scale bar corresponds to ten times the average standard error of the entries in the sample covariance matrix.

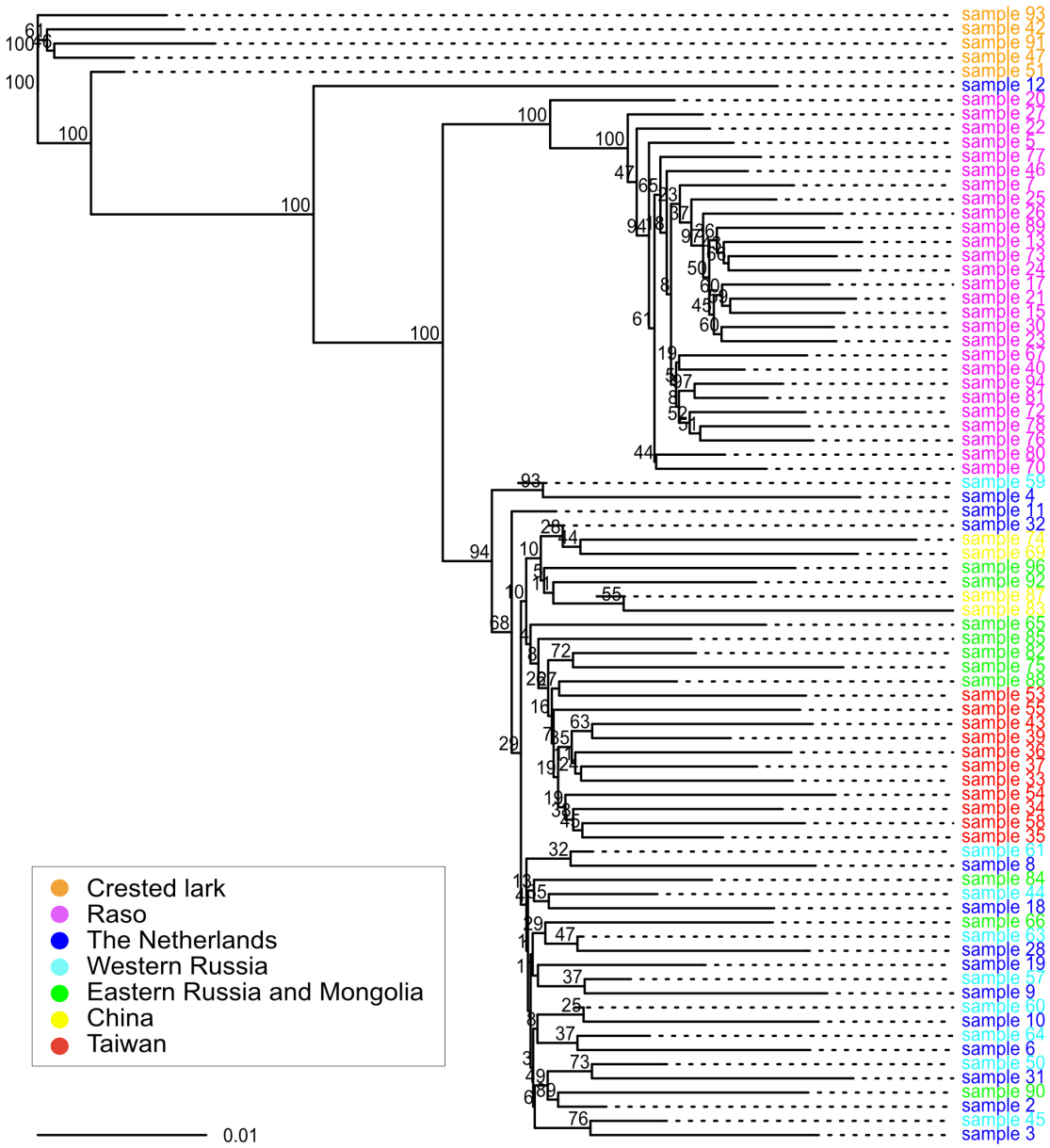


Figure 5.8 Neighbour-joining phylogenetic tree with bootstrapping values indicated at the nodes. The scale bar illustrates a 1% difference. Sample 12 is the same sample that also appears as an outlier in the Dutch population on the STRUCTURE plots (Figure 5.6), and should therefore be ignored.

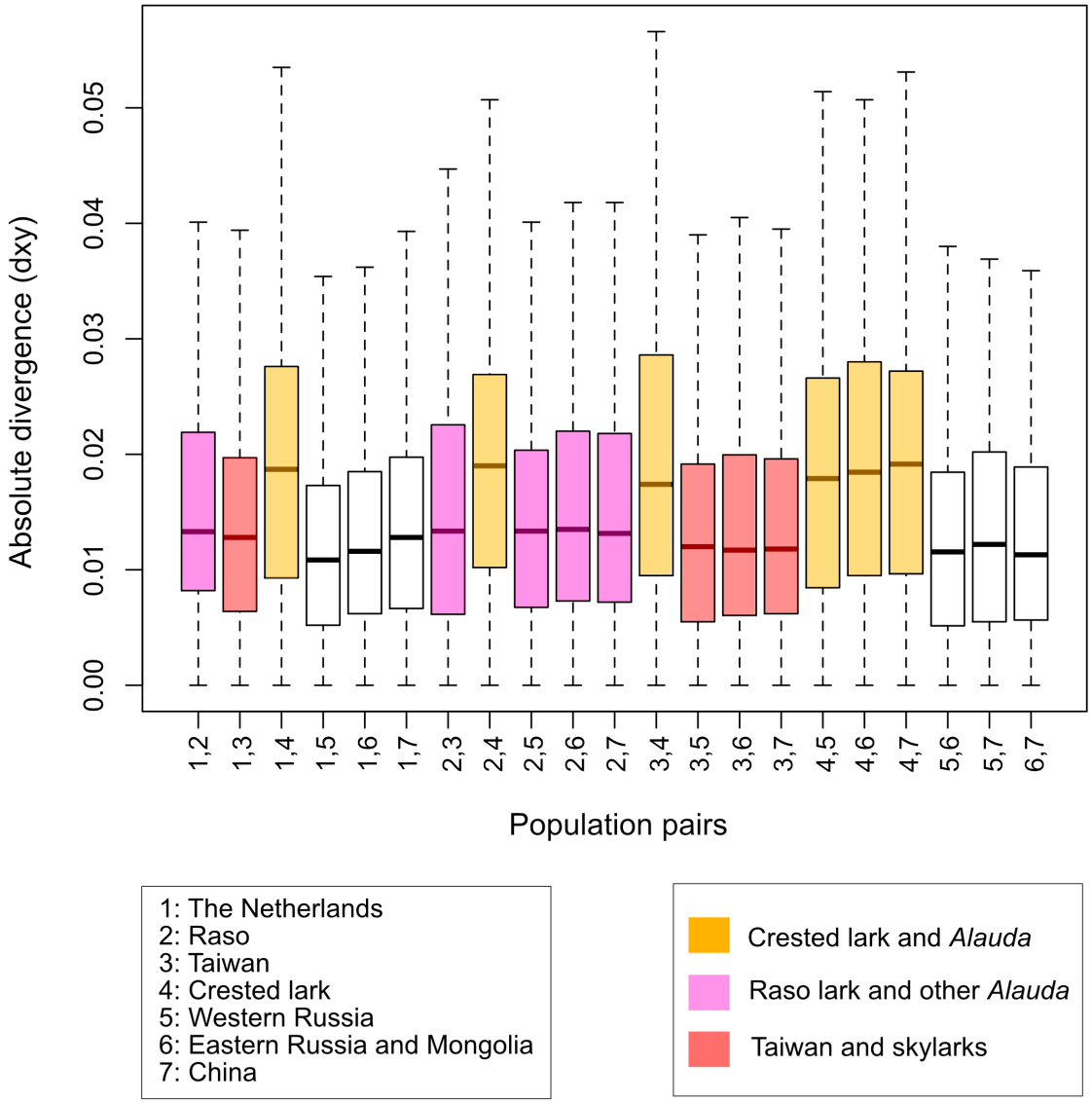


Figure 5.9 Absolute divergence (dxy) for each pair of populations. The orange box plots correspond to comparisons between the crested lark and each of the other populations, the pink box plots correspond to comparisons between the Raso lark and the other *Alauda* populations, the red box plots correspond to comparisons between the Oriental lark (Taiwan) and the skylark populations, and the white box plots correspond to comparisons between the remaining (skylark) populations. The whiskers correspond to the reasonable extremes of the data, defined as the minimum and maximum values that do not exceed a 1.5 times the interquartile range.

Discussion

This chapter presents the first whole skylark genome. Its estimated size is 1.06 billion base pairs, and the GC content of the fragment reads is 42.9%. This genome provided the opportunity to compare RADseq results based on a reference genome

alignment versus a *de novo* alignment. The *de novo* alignment resulted in more SNPs than the reference genome alignment (Table 5.2). Results of the different analyses did not differ substantially between the two methodologies. Based on this, *de novo* alignments of RAD sequences in Stacks appear to be a very high-quality alternative if a reference genome is not available. Furthermore, the fact that the reference genome used was more closely related to some of the populations than others does not seem to bias results.

In all analyses, the Raso lark was clearly differentiated from the skylark and the Oriental lark, which strongly confirms its status as a separate species. This was not unexpected, based on its clearly delimitated range and its morphological characteristics, such as the lack of rufous tones in its plumage, its smaller body size and its heavier, sexually dimorphic bill (Donald 2004). On the other end of the spectrum, the skylark populations from the Netherlands and from Western Russia seem to be particularly panmictic and undifferentiated from each other based on the F_{ST} values, the STRUCTURE analysis and the NJ tree.

All analyses except the DAPC - that is, the F_{ST}, STRUCTURE, ML tree and NJ tree analyses - show the Oriental lark as one skylark population amongst others. Consequently, the status of the Oriental lark as a separate species may not be warranted. Instead, based on the STRUCTURE figures, there seems to be a west-east gradient of skylarks across Eurasia, with the Oriental lark as the eastern extreme of the gradient. In this scenario, the skylark and the Oriental lark are a single species that has started speciating on each end of the range. Alternatively, they could be two species - one in the west and one in the east - that have come into contact in the middle of the range and have interbred there. If this is the case, it is likely to be a relatively old event and is unlikely to still be happening today. If it were, one would expect much more variation between individuals from the central Eurasian populations on the STRUCTURE plots. One would also expect higher F_{ST} values between the Oriental lark and all skylark populations, since they are at one extreme of the range and hence would be expected to be the most differentiated. Instead, these F_{ST} values are not very high; in fact, they are lower than the ones between China and the other skylark populations. Furthermore, Taiwan is an island population, subject to genetic drift, which is probably why the Taiwanese individuals cluster together on the DAPC plot and the NJ tree. If I had sampled a mainland Oriental lark population, it is likely that the individuals would not have clustered together, instead interspersing with the skylark populations. While I sampled only one Oriental lark population, Taiwan is the population that is the most

"distant" from the skylark that I could have sampled: it is an island population, it lies at the Eastern extreme of the range of both species, and no skylarks are found on Taiwan. Thus, other Oriental lark populations - mainland, mid-range populations that are potentially hybridizing with the skylark - would have been even less likely to be differentiated from the skylark. Additionally, it is worth remembering that a DAPC is a tool that maximizes variation, "exaggerating" apparent differentiation between populations.

Based on these results, this chapter provides an answer to the previously unresolved *Alauda* node from Alström et al. (2013). The Raso lark and the skylark are sister species, and the Oriental lark branch should be collapsed with the skylark branch. I estimated a coalescence time of 5 million years between the Raso lark and the other members of the *Alauda* clade, and a coalescence time of 6.4 million years between the crested lark and the *Alauda* clade. Alström et al. (2013) estimated the split between the Raso lark and the other *Alauda* members to have taken place 5.5 million years ago, and the split between the crested lark and *Alauda* 8.5 million years ago.

The Oriental lark appears to be a subpopulation, or maybe a subspecies, of the skylark. This idea is not new: in his review of the larks, Meinertzhagen (1951) considered the Oriental lark to be a subspecies of the skylark. Indeed, in light of this new molecular evidence and of the substantial amount of intraspecific morphological variation in both the skylark and the Oriental lark, their interspecific morphological differentiation seems quite minor and maybe insufficient to justify their classification as separate species. At a higher taxonomic level, Harrison (1966) suggested that many lark genera should also be collapsed.

In birds more generally, approaches to species delimitation have varied over time, and some major historical trends can be discerned. At the time of publication of Sharpe's monumental "Hand-list of the genera and species of birds" in 1909, 18,939 bird species were recognized, based on Linnaeus's binomial system. Around that time, certain ornithologists, most notably two Americans, Elliott Coues and Robert Ridgway, and two Germans, Ernst Hartert and Karl Jordan, started using trinomials to collapse poorly morphologically differentiated species (Coues 1882; Ridgway 1919; Hartert 1922; del Hoyo and Collar 2014). In this fashion, many species were re-classified as subspecies; for example, *Otis dybowskii* became *Otis tarda dybowskii* (Taczanowski 1874). In the thirty years that followed Sharpe's death in 1910, this process based on the trinomial system called "lumping" reduced the number of bird species by more than 10,000. Most of these decisions have been accepted by the ornithology community, with

a few changes being made in more recent times. These more recent changes resulted in a slight renewed increase in the number of species: in total, 8,616 species were recognized in 1946, 9,021 in 1980 and 9,672 in 1990. Today, the total is close to 10,000 (del Hoyo and Collar 2014).

Part of this "recent" increase in the number of bird species can be explained by readjustments following the "lumping" period (Ridgway 1923) and advances in molecular methods, which can detect discrete evolutionary units at a finer scale (Barrowclough et al. 2016). This trend is not frequently challenged in the literature, "by virtue of a domino effect involving one uncritical acceptance after the other, each exerting an ever-increasing pressure to conform" (del Hoyo and Collar 2014). In conclusion, in addition to solving the phylogenetic relationships between members of the *Alauda* clade, the research in this chapter provides a useful molecular-based counter-example to this trend.

Chapter 6: Genomic comparison of an island endemic and its widespread sister species

Raso lark male in front of a Raso lark burrow. © Edwin Winkel

Abstract

This chapter aims to compare the population genetics of the Raso lark, a Critically Endangered island endemic, with those of the skylark, its widespread continental closest relative. In particular, this chapter estimates the genetic diversity of the Raso lark - an important factor to take into consideration for conservation planning - and investigates the drivers contributing to the observed diversity. The Raso lark was found to have unexpectedly high levels of nucleotide diversity (only 18-35% lower than the skylark). A simulation shows that the population contraction through which the species passed was recent enough for most of the nucleotide diversity (80-100%, depending of the severity of the contraction) to be retained in the present-day population. Furthermore, in 15 out of 27 individuals, 16 percent of the genome had levels of heterozygosity on average 6.6 times higher than elsewhere on the genome, likely due to suppressed recombination and the existence of a neo-sex chromosome in larks. However, individual heterozygosity on the recombining parts of the genome was low, 81% of pair-wise

comparisons between sampled individuals indicated third-degree kinship or higher, and the Raso lark showed very high levels of linkage disequilibrium compared to the skylark and other bird species. All in all, despite high levels of nucleotide diversity, other genetic signatures clearly show that the Raso lark underwent a severe population contraction several centuries ago.

Introduction

Chapter 5 established the clear genetic differentiation between the Raso lark and its two widespread, continental close relatives, the skylark and the Oriental lark. The goal of the present chapter is to delve deeper into this topic and, focusing on the Raso lark, gain a better understanding of the species' demographic history. What signatures have demographic factors such as small population size, population contraction, bottlenecks, restricted range and geographic isolation left on the Raso lark's genome?

Perhaps the most crucial component of this question is to compare the levels of genetic diversity present in the Critically Endangered and island endemic Raso lark with those of its widespread, continental relatives. Indeed, the maintenance of genetic diversity is crucial for fitness and survival, at the individual, population and species level (Frankham, Ballou and Briscoe 2002). Genetic diversity increases the viability of recently translocated populations. It also favours ecosystem recovery and resilience in the face of disruptive events such as extreme weather episodes. Finally, it helps species respond to environmental changes such as climate change, to which Cape Verde is particularly vulnerable (Keller and Waller 2002; Reusch et al. 2005; Hughes et al. 2008; Wright et al. 2009; Johnson et al. 2010; Ministry of Environment Housing and Territory Planning of Cape Verde 2011; Tollington et al. 2013; Nair 2014; Romiguier et al. 2014). Genetic diversity in an island population is particularly important. There is evidence that species on smaller landmasses (islands) have lower rates of molecular evolution than species on larger landmasses (Wright et al. 2009; Tollington et al. 2013). This suggests that confining species to small refugia reduces the rate of microevolution, which could limit the species' ability to adapt to unprecedented environmental changes (Wright et al. 2009; Tollington et al. 2013).

While this is significant for the current Raso lark population on Raso, it is yet more relevant for plans to translocate birds to Santa Luzia (see Chapter 1). While this translocation project is of utmost importance for the conservation of the species, it could in practice constitute another bottleneck, if the founder population is too small. The

success of previous reintroduction programs for other species has been highly variable, with inbreeding in small founder populations and the resulting vulnerability to disease often being cited as a cause of failure (Frankham, Ballou and Briscoe 2002; Brekke et al. 2010). Indeed, the first aim of conservation managers is often to quickly increase population numbers. In doing so they often overlook the long-term genetic effects of reintroductions and post-release management (Tollington et al. 2013). However, a review of genetic diversity in introduced species found that, although loss of diversity compared to the source population is common after introductions, it is not universal. In a few rare cases - two plants and one fish species - there have even been large increases in diversity, because founder individuals stemmed from different source populations (Dlugosch and Parker 2008). When planning a translocation, the genetic diversity of the source population, in this case Raso, will need to be taken into account.

Genetic diversity is often seen as resulting from the interaction between mutation rate and historical events, whether demographic or adaptive: population structure, population bottlenecks, selective sweeps and/or ecological disturbances (Romiguier et al. 2014). The extent of the impact of species biology and ecology is still largely unknown, but Romiguier et al. (2014) demonstrated that a species' genetic diversity can be predicted from its ecological strategy. They calculated genome-wide synonymous nucleotide diversity (π_s) for 76 non-model animal species by sequencing the transcriptome of two to ten individuals in each species, and compared that with the species' life history traits (adult size, body mass, maximum longevity, adult dispersion ability, fecundity and propagule size). All these traits were significantly related to π_s. More specifically, species traits related to parental investment predicted genetic diversity: long-lived or low-fecundity species with brooding ability (K-strategy species) tend to be genetically less diverse than short-lived or highly fecund ones (r-strategy species). The theoretical explanation for this is that, since neutral genetic diversity increases with effective population size Ne, π_s is inversely linked to biological traits related to abundance, such as body size or fecundity. Romiguier et al. (2014) suggest that, "because K-strategy species have been selected for survival and the optimization of offspring quality in complex, stable environments, […] they might experience fewer occasional disturbances (or be less sensitive to them), thus ensuring the long-term viability of even small populations. In contrast, only species with a large population-carrying capacity could sustain the "riskier" r-strategy in the long term, thus buffering the frequent bottlenecks experienced in the context of high environmental sensitivity."

The Raso lark, with its current small population size and risk of inbreeding,

known historical population contraction, bottlenecks and sex ratio biases, has undergone at least some of these historical events that put species at risk of lower genetic diversity (Gossmann et al. 2012). Furthermore, as seen in Chapters 3 and 4, the Raso lark is closer to the K end of the r/K strategy gradient than its closest relatives. This is an additional argument to predict lower genetic diversity in the Raso lark than in the skylark and the Oriental lark. Although based on these arguments the Raso lark's genetic diversity might be predicted to be low, a certain amount of genetic variation could still be preserved in the species from a time when it was much more abundant, if that time was relatively recent. As a matter of fact, the Raso lark was much more widespread in Cape Verde until the arrival of the Portuguese in 1462 (see Chapter 1). Furthermore, a neo sex chromosome found in some larks species, including the Raso lark, has been suggested as a mechanism to maintain heterozygosity in the genome (Brooke at al. 2010).

The aim of this chapter is to compare the population genetics of the Raso lark, a Critically Endangered island endemic, with those of the skylark, its widespread continental closest relative. In particular, this chapter estimates the genetic diversity of the Raso lark and investigates the drivers contributing to that with, as the end goal, to produce results that are useful for the planned reintroduction onto Santa Luzia Island, as will be discussed in Chapter 7.

Methods

Sampling, laboratory work, sequencing, sequence alignment and sequence processing were done as described in Chapter 5. Based on the results in Chapter 5 that showed very little to no differentiation between skylark populations, comparisons between the Raso lark and the skylark are preferentially done using the skylark population from the Netherlands, since that is the one with the largest sample size. Furthermore, using a single geographic locality for the skylark is the most appropriate approach to comparisons with the Raso lark, since it prevents biasing the analyses with the detection of genetic diversity present due to isolation by distance in the skylark, or the larger geographic range of the skylark. The Taiwanese population is also used as an additional reference point, since it is an island population like the Raso lark, albeit larger. For certain analyses, other lark populations are used for comparison, for reasons explained in the methodology paragraph of that specific analysis.

Population genetic diversity

To estimate levels of genetic diversity in the Raso lark, the relevant statistics (nucleotide diversity π , observed homozygosity, observed heterozygosity, expected homozygosity, expected heterozygosity) were calculated in Stacks 1.35 (Catchen et al. 2013) based on datasets 1 and 7 (Table 5.2). I then calculated the relative difference in π between Raso and each of the other populations, according to the formula:

$$\frac{x - y}{y} \times 100$$

where *y* corresponds to Raso and *x* to the other population. I also calculated nucleotide diversity after excluding samples 4, 12 and 20, three outliers with increased individual heterozygosity, possibly due to laboratory or sequencing contamination, but this had no notable effect on the results.

The values of π obtained were used to calculate the effective population size (Ne) of the Raso lark according to standard population genetic theory and assuming neutrality, using the relation $\pi = 4Ne\mu$:

$$Ne = \frac{\pi}{4\mu}$$

where μ is the mutation rate per site per generation. I used two different estimates of the mutation rate: 1.5×10^{-9} per site per year in birds (Ellegren et al. 2007) and 2.3×10^{-9} per site per year in collared flycatchers *Ficedula albicollis* (Smeds, Qvarnström and Ellegren 2016). To obtain a mutation rate per generation for the Raso lark, I calculated the generation time of the Raso lark, defined as the mean age of the parents of the current cohort, as follows, using the values of annual mortality obtained in Chapter 3:

$$generation\ time = \frac{1}{mean\ annual\ mortality} + age\ at\ first\ breeding$$

Investigating the reasons behind high nucleotide diversity

I show below that values for π_{Raso} were higher than expected for such a small population. I next investigated the two most likely explanations for this.

Hypothesis A: population demographic history. The Raso lark population was much larger in the past and has undergone a very recent population contraction that has not yet resulted in a large decrease in π, in accordance with this formula from population genetics theory (Frankham, Ballou and Briscoe 2002):

$$H_t = H_0 \times \left(1 - \frac{1}{2N}\right)^t$$

where H_0 is the initial heterozygosity in the population, H_t the heterozygosity in the population at a time t and N the population size. To roughly estimate the Raso lark population size in Cape Verde before the arrival of the Portuguese in the 15th century, I summed the area of the four islands where the Raso lark used to be found according to the fossil record - Raso, Santa Luzia, São Vicente and Santo Antão (Mateo et al. 2009) - and extrapolated the pre-Portuguese population size from the current Raso lark population size. This pan-Cape Verde lark population is very likely to have been panmictic: the aforementioned islands are situated close to each other, the larks are strong fliers as evidenced by the males' impressive courtship displays despite the strong winds of Raso, and Raso larks have occasionally been seen on other islands in recent years, including a sighting on São Nicolau (Hazevoet 2012). In collaboration with Simon Martin, I then more formally investigated the role of this past demographic event in explaining the high observed nucleotide diversity in the Raso lark, by running a simulation to examine loss of genetic diversity over time. We used ms (Hudson 2002) to simulate populations experiencing a recent contraction that lasts for a certain number of generations to the present. The ancestral population size was set at 100,000 diploids, and three different contraction effective population sizes were considered: 10,000, 1000 and 100. The simulation was based on 28 diploid samples. π was computed for 50,000 loci of 100bp each, with a θ value of 2%.

Hypothesis B: a neo sex chromosome. The second most likely explanation for the high nucleotide diversity in the Raso lark, one that does not rule out Hypothesis A, is derived from a study by Brooke et al. (2010), which shows that the Raso lark has unusually large sex chromosomes, multiple nuclear markers that appear to have become sex-linked, and allele distributions indicative of homologous W and Z alleles. Based on these results, the Raso lark appears to have experienced a translocation or fusion of genetic material from the autosomes to both sex chromosomes, as do other lark species, including the skylark (Bulatova 1973; Pala 2012). The addition of autosomal genes to the sex chromosomes might be a mechanism by which some level of heterozygosity could be maintained in the Raso lark, both at an individual and at a population level: the two sex chromosomes maintain two separate lineages for all genes that reside on them. Sex chromosomes are also subject to genetic drift - in fact, to a higher degree than autosomes, if they do not undergo strong balancing selection, due to lower Ne.

However, even though W and Z tend to each rapidly become monomorphic in small populations, because each sex chromosome carries different alleles, overall greater diversity is maintained, at least until the W chromosome degeneration becomes more advanced (Brooke et al. 2010).

Brooke et al. (2010) suggest that the neo sex chromosome appeared recently. This is based on two arguments: first, that the karyotypes produced by Bulatova (1973) showed enlarged sex chromosomes "in some but not all larks;" second, that little time can have elapsed since the fusion, since the microsatellite alleles exist on both the W and Z chromosomes, and W chromosomes generally degrade rapidly. Pala et al. (2012), on the contrary, found evidence that the neo sex chromosome arose at the base of the Sylvioidea, 42.2 million years ago. Today, the Pala et al. (2012) hypothesis seems more likely, because a recent study showed that many species of birds degenerate very slowly in their W chromosomes (Wang et al. 2014), and a careful reading of the Bulatova (1973) paper indicates that the author karyotyped four lark species, found enlarged sex chromosomes in three (the red-capped lark *Calandrella cinerea*, the bimaculated lark *Melanocorypha bimaculata* and the horned lark *Eremophila alpestris*), and inconclusive results - *not* small sex chromosomes - for the fourth (the crested lark).

To test whether individual heterozygosity varies across the genome as predicted by the neo sex chromosome theory, in collaboration with Simon Martin, a python script (*popgenWindows.py* available at https://github.com/simonhmartin/genomics_general/) was used to calculate individual heterozygosity per 100kb window for the four different lark species that I sequenced in Chapter 5 (Raso lark, skylark, Oriental lark and crested lark). The microsatellite results from Brooke et al. (2010) - in which, for all markers, females were always heterozygous and males could be either heterozygous or homozygous (Tables 4.2 and 4.S1) - predict that, in this study, males can have locally elevated heterozygosity (called "heterozygote individuals") or not (called "homozygote individuals"), while females would all fall in the second category. As for the timing of the emergence of the neo sex chromosome, if it was a recent event as suggested by Brooke et al. (2010), the crested larks in this analysis will all be homozygotes. If the fusion was basal to all larks, as suggested by Pala et al. (2012), all the lark species sampled in this study, including the crested lark, could have both homozygote and heterozygote individuals.

This analysis was performed in collaboration with Simon Martin, on dataset 1, but with relaxed dataset completeness requirements for site inclusion: the read depth per individual was lowered to 5, in order to maximize the number of sites per window, since

this script works on 100kb windows. Windows are rejected if they have fewer than 100 sites meeting the site inclusion requirements. There was no requirement for a site to be present in a certain number of individuals to be included, since the focus of this analysis is the individual, not the population. Scaffold 2 looked like a possible mis-assembly (one part had low heterozygosity and the other had high heterozygosity) and hence was excluded. One Raso lark sample, number 20, looks very different from all other Raso larks (very high individual heterozygosity across the genome), possibly due to laboratory or sequencing contamination. As such, it was not included in these results. Because the sex of individuals needs to be known for this analysis, I was not able to use the skylark population with the largest sample size, The Netherlands, to represent the species, since these birds were bled as chicks and were not sexed (Table 5.S1). Instead, I used the skylark population with the second largest sample size, Eastern Russia and Mongolia (Figure 5.3).

Relatedness

Since levels of kinship and hence inbreeding could be very high in a population as small as the Raso lark, I ran the program KING (Manichaikul et al. 2010) with option --*homo* to calculate relatedness between Raso lark individuals. KING was run on the RADseq data processed with *process_radtags* as described in Chapter 5, aligned *de novo* with parameters -m 3 -M 3 -n 5 and filtered further with *populations* parameters set at -m 10 -p 2 -r 0.5. Additional filtering was performed with PLINK 1.07 (Purcell et al. 2007), keeping only SNPs with a minor allele frequency greater or equal to 0.1, with a genotyping rate of 40% or higher, and passing a Hardy-Weinberg Equilibrium test at the 0.05 significance level. I used a *de novo* alignment for this analysis because the values obtained with a reference genome alignment were much too high to be realistic, for example indicating first-degree and even clonal relationships between crested larks sampled in different parts of Saudi Arabia. The *de novo* dataset, on the contrary, gave plausible results, and correctly identified the known pair of siblings amongst the Dutch skylark samples.

Linkage disequilibrium

Linkage disequilibrium, also referred to as LD or r^2, is the non-random association of alleles at different loci. It can provide a good insight into the population

genetic forces that shape a genome (Slatkin 2008). I calculated LD in order to potentially detect an effect of the population contraction on the Raso lark genome. Using formulas outlined in Wakeley (2008), I calculated r^2 between sites for loci in dataset 1 with more than one SNP, to determine the level of non-independence of the SNPs. For these calculations, I randomly chose 11 individuals from each population because unequal sample sizes can bias the calculations of r^2. Calculations were performed with a python script written by Allison Shultz (*Pairwise_linkage_disequilibrium.py* available at https:// github.com/ajshultz/Rad/).

F_{ST} outliers

Given the stark contrast between the life strategy and environment of the Raso lark and the skylark, I looked for genomic sites that might be under specific selection in *Alauda razae*. For this, I used BayeScan 2.1 (Foll and Gaggiotti 2008) to detect F_{ST} outliers amongst Raso lark and Dutch skylark SNPs. Because BayeScan can be sensitive to loci with low minor allele frequencies, I filtered dataset 1 to retain only loci with a minor allele frequency greater than 0.1. I used a reference genome alignment since that is recommended over the use of a *de novo* alignment in order to avoid false positives and increase the efficiency of the detection of loci underlying adaptive change (Nadeau et al. 2014). BayeScan was performed with a burn-in of 50,000 iterations and 100,000 generations of data collection in dataset 1. Then, I took the 200 base pairs flanking the putative SNP under selection or, when the SNP was less than 100 base pairs away from the start or the end of the scaffold, until scaffold start or end positions. This sequence was then blasted (Altschul et al. 1990) against the best annotated passerine genome, the zebra finch genome (Warren et al. 2010), in order to gain an insight into its biological function.

Results

Population genetic diversity

I found $\pi_{Raso} = 0.0114$ for dataset 1 and $\pi_{Raso} = 0.0070$ for dataset 7. While the absolute values of π varied between datasets, the Raso lark consistently had the lowest nucleotide diversity amongst all populations studied, and this result was consistent between read alignment methods. For dataset 1, the relative difference in π between Raso and the Netherlands - the population with the highest nucleotide diversity - was 35%. The relative difference in π between Raso and the population with the next lowest nucleotide diversity, Western Russia, was 18% (Table 6.1).

The same pattern is apparent when looking at observed heterozygosity and homozygosity: the Raso lark has the lowest proportion of heterozygous individuals in its populations and the highest proportion of homozygotes amongst all populations (Table 6.1). The proportions of homozygotes and heterozygotes expected under Hardy-Weinberg equilibrium are included in Table 6.1 for comparison ("expected heterozygosity" and "expected homozygosity").

Based on the values of π, on the estimates of the mutation rate in birds from Ellegren et al. (2007) and Smeds, Qvarnström and Ellegren (2016), and on a calculated generation time of 6.5 years, I estimated an effective population size of the Raso lark ranging from 116,430 to 290,744 individuals, which is much higher than the census population size (≈ 1000). For the Dutch skylark population, based on the same mutation rates and a calculated generation time of 3 years, I estimated an effective population size ranging from 373,188 to 855,556 individuals.

Table 6.1 Genetic diversity statistics (π, observed and expected heterozygosity, observed and expected homozygosity) for each population based on datasets 1 (in black) and 7 (in blue).

Population	Raso	The Netherlands	Western Russia	Eastern Russia and Mongolia	China	Taiwan
π	0.0114 0.0070	0.0154 0.0103	0.0134 0.0082	0.0143 0.0098	0.0147 0.0116	0.0143 0.0090
Relative difference in π with Raso (%)	0 0	35 47	18 17	25 40	29 66	25 29
Observed heterozygosity	0.0179 0.0103	0.0200 0.0122	0.0185 0.0111	0.0188 0.0124	0.0187 0.0136	0.0204 0.0120
Observed homozygosity	0.9821 0.9897	0.9800 0.9878	0.9815 0.9889	0.9812 0.9812	0.9813 0.9864	0.9796 0.9880
Expected heterozygosity	0.0111 0.0068	0.0146 0.0096	0.0124 0.0075	0.0133 0.0090	0.0119 0.0091	0.0134 0.0084
Expected homozygosity	0.9889 0.9932	0.9884 0.9904	0.9876 0.9925	0.9867 0.9910	0.9881 0.9909	0.9866 0.9916

Investigating the reasons behind high nucleotide diversity

Hypothesis A: population demographic history. Adding up the areas of Raso (7km^2), Santa Luzia (35km^2), São Vicente (227km^2) and Santo Antão (779km^2) (Mateo et al. 2009), I obtained a total area of 1,048km^2, which corresponds to 150 times the area of Raso. Since the current population on Raso is ≈ 1000 individuals, we can extrapolate to a pre-Portuguese population size of 150,000 larks. The ms simulation indicates that nucleotide diversity remains high in populations for a large number of generations after a population contraction (Figure 6.1). The Portuguese settled Cape Verde in 1462, which roughly corresponds to 85 Raso lark generations ago. A population undergoing a reduction from 100,000 to 1000 individuals maintains virtually all of its nucleotide diversity after that amount of time. Even in the case of a more severe reduction, with a population dropping to 100 individuals (as the Raso lark

population has probably done several times since 1500), more than 80% of the diversity is maintained (Figure 6.1).

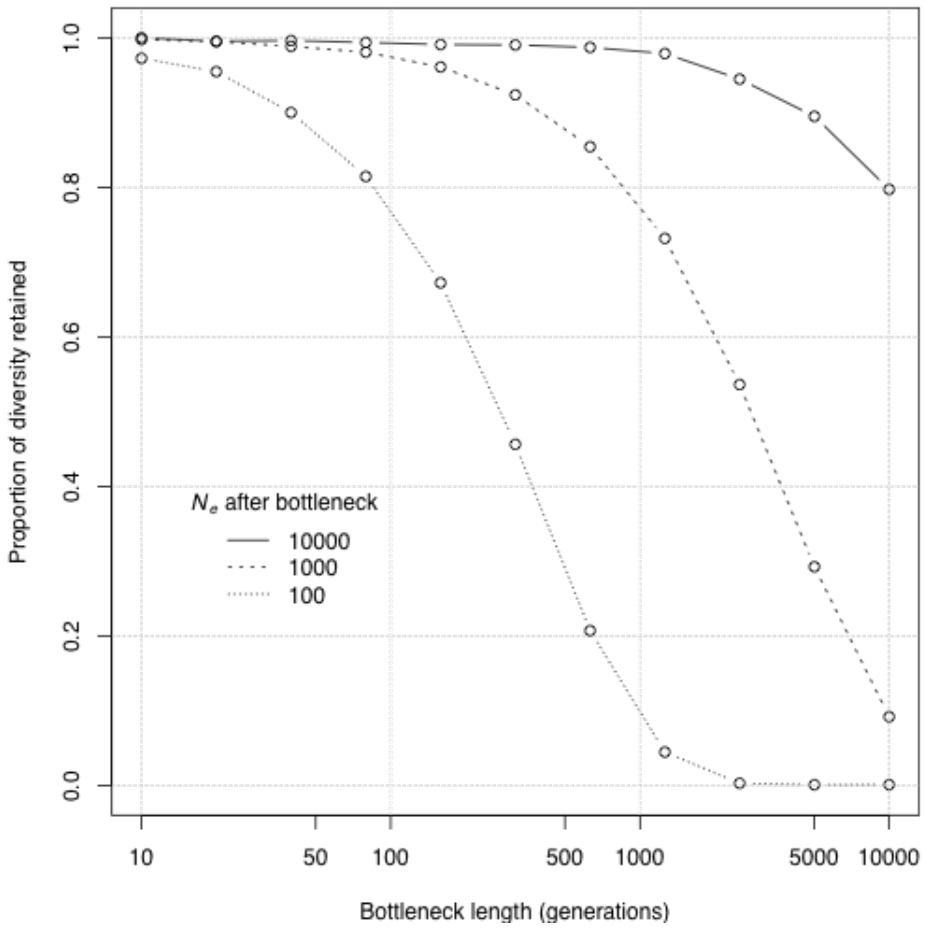


Figure 6.1 Plot based on simulated data showing the ratio of π in the reduced population to π computed for a constant population.

Hypothesis B: a neo sex chromosome. In 15 out of 27 Raso lark samples, heterozygosity varies dramatically across the genome: while "baseline" heterozygosity is very low, some scaffolds have elevated levels of heterozygosity (Figure 6.2). These scaffolds probably have elevated heterozygosity due to suppressed recombination, possibly due to linkage to the sex chromosomes.

For further calculations, conservative sets of scaffolds that putatively experience Suppressed Recombination ("SR") or Normal Recombination ("NR") were identified based on heterozygosity in the Raso lark samples, as follows. First, all scaffolds shorter than 1Mb were discarded, as most smaller scaffolds lack sufficient numbers of genotyped sites to be confidently assigned to one of the two groups. This left 315 scaffolds. Of these, 265 were confidently designated as NR scaffolds, with heterozygosity < 0.005 in all samples, and 40 were designated as SR scaffolds, with

heterozygosity > 0.01 in at least one sample. Ten scaffolds were discarded as ambiguous, having intermediate diversity in at least one sample. The 305 scaffolds assigned to the two groups represents 61% of the genome, or 631,562,720 base pairs. Of this, 16% is made up of SR scaffolds and 84% of NR scaffolds. Mean heterozygosity across Raso lark samples at the SR scaffolds is 0.008, while mean heterozygosity at the NR scaffolds is 0.0012 - that is, over six times lower.

Based on this discovery, I re-calculated π using only the NR scaffolds in dataset 1, and found $\pi_{Raso} = 0.0082$ and $\pi_{The\ Netherlands} = 0.0165$. The relative difference between these two values is 67%. This revised value of π_{Raso}, the mutation rate in birds from Ellegren et al. (2007) and a generation time of 6.5 years give us an effective population size of the Raso lark of 209,131 individuals, which is lower than the original estimate that included all scaffolds (290,744 individuals), but is still much higher than the census population size (≈ 1000).

Of the 15 Raso larks with the pattern of locally elevated heterozygosity, called heterozygotes, five are male, nine are females, and one is of unknown sex (Figure 6.3). Of the 12 Raso larks *not* showing this pattern, 11 are male, and only one is female (sample 26; Figure 6.3). Elevated heterozygosity on certain scaffolds is also present in individuals from other lark species (Figures 6.2 and 6.3). However, these heterozygotes are much harder to detect than in the Raso lark, because these species are more genetically diverse than the Raso lark: the difference in heterozygosity between regions with and without suppressed recombination is minimal, especially in the skylark (Figures 6.2 and 6.3). These other species of larks also have a pattern of males being either homozygotes or heterozygotes, and females being heterozygotes (Figure 6.3).

Relatedness

The mean pair-wise relatedness of all sampled Raso lark individuals was 0.082, which falls within KING's third-degree kinship bracket. Of all 464 possible pairs of sampled individuals, 1% were first-degree relationships, 46% were second-degree relationships, 34% were third-degree relationships and 19% were related to a lesser degree.

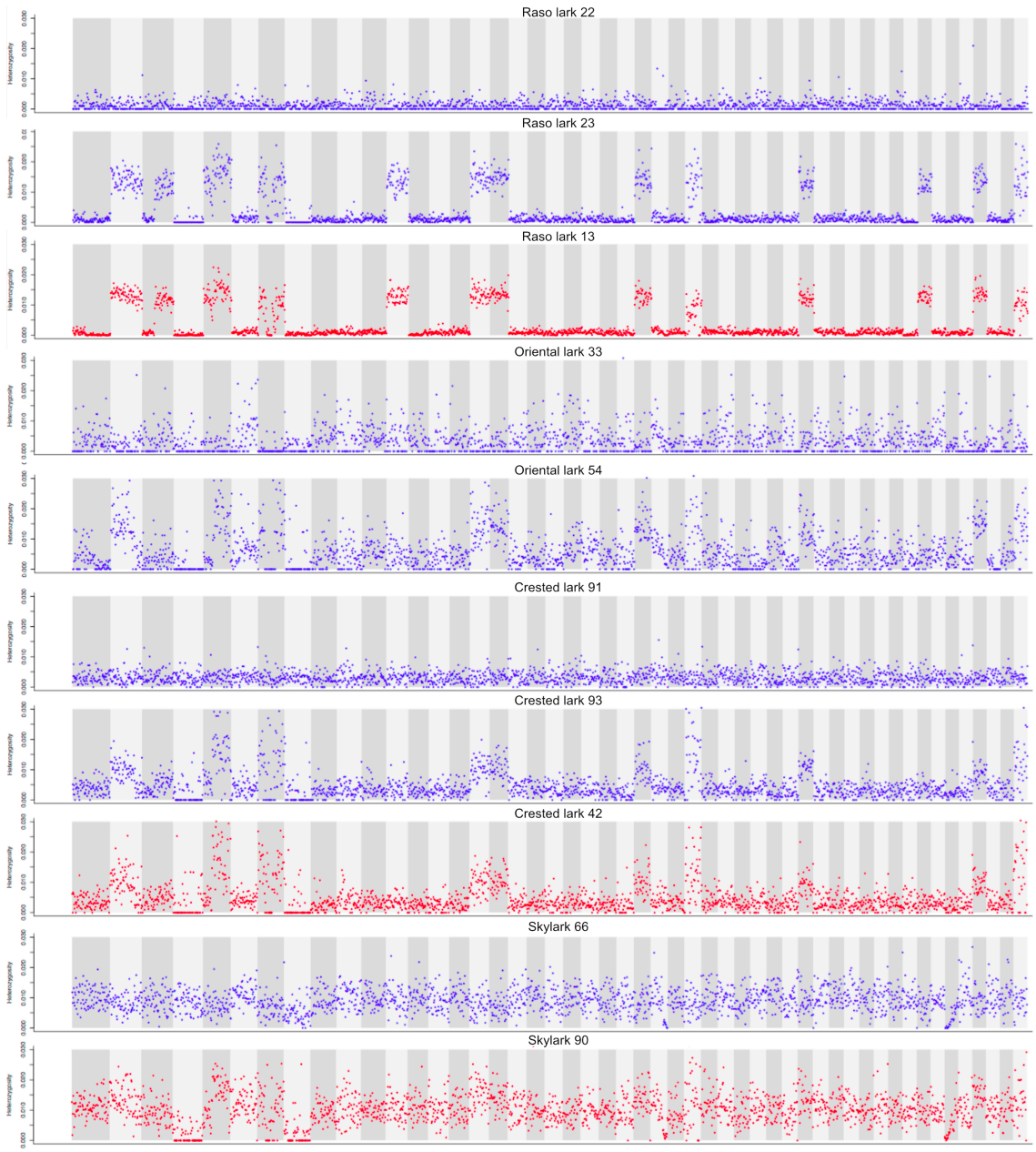


Figure 6.2 Individual heterozygosity in non-overlapping windows per scaffold for the 50 largest scaffolds in 10 representative larks: a homozygous Raso lark male (sample 22), a heterozygous Raso lark male (sample 23), a heterozygous Raso lark female (sample 13), a homozygous Oriental lark male (sample 33), and heterozygous Oriental lark male (sample 54), a homozygous crested lark male (sample 91), a heterozygous crested lark male (sample 93), a heterozygous crested lark female (sample 42), a skylark male (sample 66) and a skylark female (sample 90). Males are shown in blue and females in red.

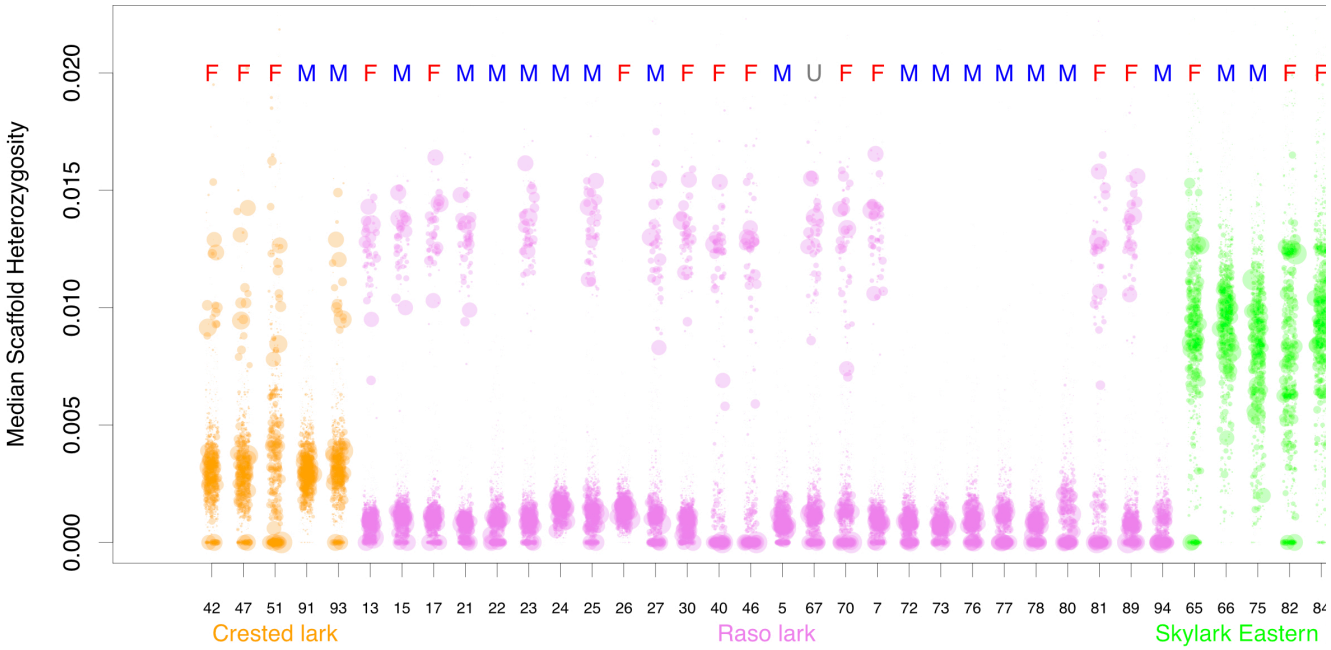


Figure 6.3 Median scaffold heterozygosity for all crested larks (orange), Raso larks (violet), skylarks (green) and Oriental larks (red) in this analysis.
Each individual's sex is recorded as either female ("F"), male ("M") or unknown ("U").

Linkage disequilibrium

For all three populations (Raso, Taiwan, The Netherlands), linkage disequilibrium among sites within loci declined gradually with distance along the sequence. The far right side of the graph is characterized by an increase in uncertainty - a result of limited data due to the short read length (Figure 6.4). Sites close together (<10 bp) exhibited r^2 values averaging 0.65 for the Raso lark and 0.47 for the other *Alauda* populations. Sites further away (> 20 bp) exhibit values closer to 0.45 for the Raso lark and 0.30 for the other larks. As such, LD is much higher in the Raso lark than in the skylark and the Oriental lark (Figure 6.4).

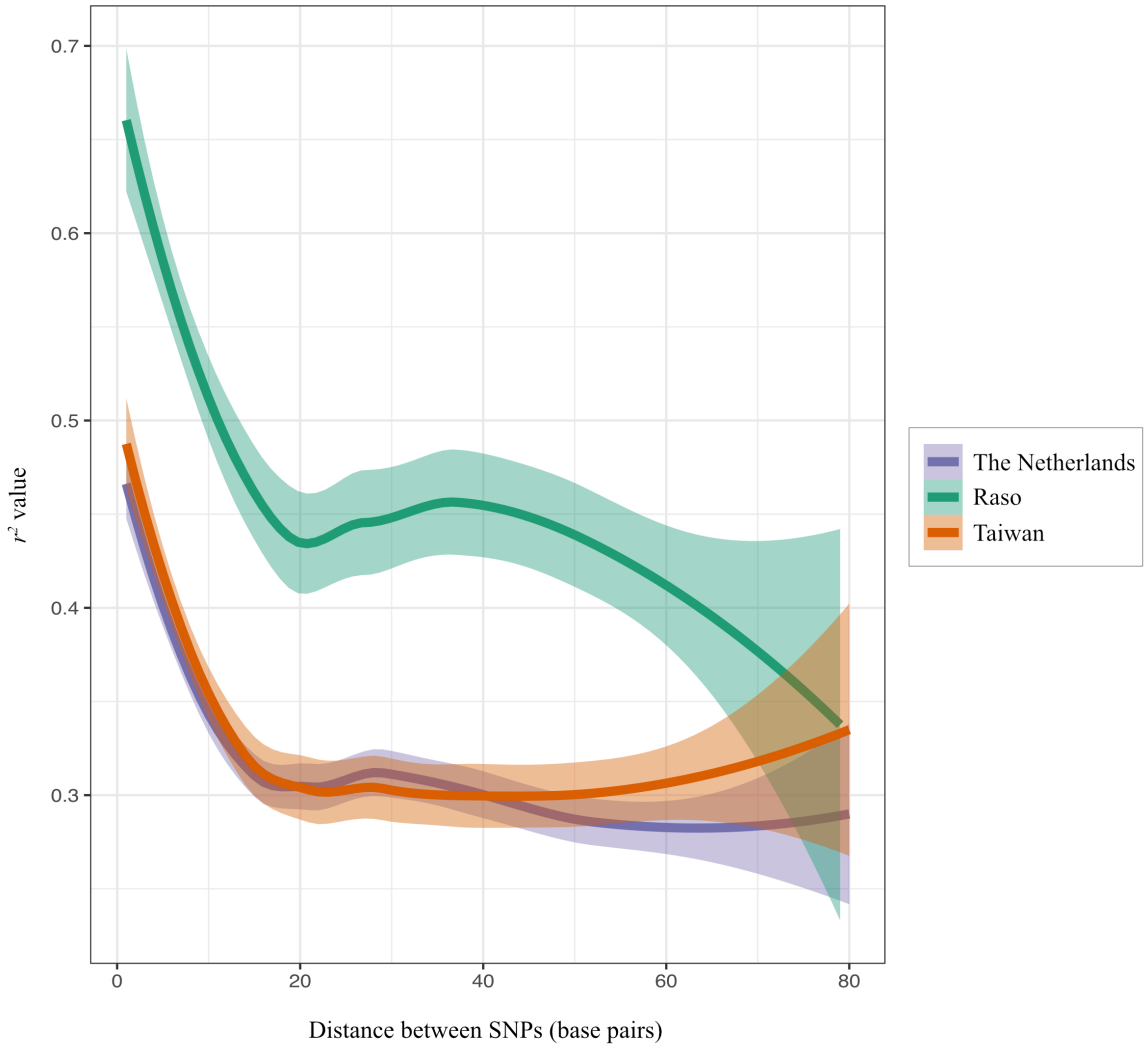


Figure 6.4 Linkage disequilibrium as measured by r^2 (Wakeley 2008) between pairs of SNPs found on the same RADseq locus. For 11 randomly chosen individuals from each population, the LOESS smoothed r^2 values (solid lines) plus 95% SE confidence intervals are shown for each SNP distance.

F_{ST} outliers

BayeScan detected one F_{ST} outlier (SNP coordinate 1059539), and the corresponding sequence matched with the zebra finch genome with an E value of 1E-64. The closest annotated gene is situated 9,430bp away from the sequence, and codes for the E3 ubiquitin-protein ligase RNF144A. This is an enzyme that mediates the ubiquitination[2] and degradation of DNA-dependent protein kinase, catalytic subunit (DNA-PKcs). DNA-PKcs plays a key role in the DNA repair pathway. Depletion of RNF144A leads to an increase in DNA-PKcs and resistance to DNA damaging agents. Hence, it is involved in protecting cells suffering from repeated or grave harm to their DNA (Ho et al. 2014). It is also linked to the shortening of telomeres and ageing, since DNA-PKcs-deficient mice have been shown to age faster and have a shorter lifespan than other mice (Espejel et al. 2004).

Discussion

High nucleotide diversity in the Raso lark

In terms of absolute values, the *Alauda* larks have levels of nucleotide diversity similar to that of the zebra finch ($\pi = 0.01$), which authors qualify as "exceptionally high" (Balakrishnan and Edwards 2009). It is, for example, much higher than that of another common and wide-ranging passerine, the house finch *Haemorhous mexicanus* ($\pi \approx 0.005$; Shultz et al. 2016), or a seabird, the black-footed albatross *Phoebastria nigripes* ($\pi \approx 0.00063$; Dierickx et al. 2015). The Raso lark's nucleotide diversity was the lowest amongst all populations studied in this chapter, including the Oriental larks from Taiwan, another - albeit larger - island population. This trend was robust across all different alignment and data filtering methods. Based on dataset 1, the Raso lark's nucleotide diversity is 35% lower than that of the skylark population from The Netherlands, the population with the highest nucleotide diversity. Although this is a clear difference, it is less pronounced than expected, given the two species' respective population sizes: the Raso lark, an island endemic, has an observed population size of ≈ 1000, while the skylark, a widespread continental species that seems to be panmictic

[2] Ubiquitination means the addition of ubiquitin to a protein to change its location in the cell, influence its activity, encourage or stop its interactions with other proteins, or, in the case of RNF144A, mark it for degradation.

across most of Western Eurasia (see Chapter 5) has 1 million pairs breeding yearly in the United Kingdom alone (Donald 2004).

When using these π values to estimate the Raso lark's effective population size, this contrast between genetic data and field observations becomes even starker. It is hard to obtain precise estimates of Ne using RADseq data, and estimates can span multiple orders of magnitude depending on the data filtering methods (Dierickx et al. 2015). However, in this case, while there was variation in the Ne estimates, it was less extreme, and Ne was consistently estimated to be two orders of magnitude larger than the census population size across all alignment and data filtering methods.

Explanations for the high nucleotide diversity

There are four possible explanations for this surprising result. The first one is methodological. Nucleotide diversity estimates can vary depending on the genetic marker and the sequencing technique used (Freeland 2005; Dierickx et al. 2015). However, while this means that absolute values of Ne and comparisons with other studies need to be treated with caution, the present study was carefully designed to include multiple other populations as reference points with which to compare the Raso lark's nucleotide diversity. Besides the absolute value of π_{Raso}, the fact remains that, based on dataset 1, it is only 18-35% lower than $\pi_{skylark}$. Furthermore, a previous study on the Raso lark using microsatellite markers found that, out of 21 markers that were successfully amplified, seven were polymorphic. To cite the authors, "to find any appreciable microsatellite polymorphism in the Raso lark is rather unexpected, given the long-term small size of the population" (Brooke et al. 2010). This evidence based on a different type of genetic marker corroborates the high nucleotide diversity found in this chapter.

The second possible explanation for the Raso lark's high nucleotide diversity is recent or current gene introgression from another species - the most likely candidate being the Raso lark's closest relative, the skylark. However, there have never been any reported sightings of skylarks in Cape Verde (Hazevoet 1995; Hazevoet 2014) and the two species' ranges do not overlap at all. The nearest skylark population, in Morocco, is located 2,200km away from Raso (Table 5.1). Furthermore, Chapter 5 shows that the Raso lark and the skylark are clearly distinct species. As such, this hypothesis seems highly unlikely, and was not explored in this study.

The third explanation is the first one that was tested in this chapter (Hypothesis

A). Could the Raso lark population before the Portuguese settled Cape Verde have been large enough, and the subsequent population contraction recent enough, for levels of genetic diversity to still be so high today? A rough estimate of past population size based on the fossil record and the area of other Cape Verdean islands yielded an affirmative answer to the first part of this question. A simulation in ms demonstrated that the Raso lark's population contraction was recent enough that the species would have conserved 80-100% of its nucleotide diversity, supporting this third explanation for the high values of π_{Raso}.

The fourth hypothesis was the second one tested in this chapter (Hypothesis B): do certain parts of the Raso lark genome have elevated levels of heterozygosity due to a neo sex chromosome that maintains two different lineages for numerous genes, as suggested by Brooke et al. (2010)? This chapter does confirm that, in certain individuals, some parts of the genome (16%) are more diverse than others, with levels of heterozygosity that are on average 6.6 times higher, and that this could indeed be a mechanism that contributes to the unexpectedly high π values found in the Raso lark. Indeed, re-calculating π based only on the non-SR scaffolds leads to a value for π_{Raso} that is 67% lower than $\pi_{The\ Netherlands}$, instead of 35% lower if all scaffolds are taken into account.

This pattern was present in other lark species as well, including specifically in the crested lark, the species for which the search for enlarged sex chromosomes was inconclusive (Bulatova 1973). The fact that males can be either homozygous or heterozygous while females are overwhelmingly heterozygous almost, but not perfectly, fits the microsatellite-based results from Brooke et al. (2010): one female, Raso lark number 26, was identified as homozygous in the present study. It is the only homozygous female out of 12 homozygous Raso larks, out of 10 female Raso larks, and out of 13 female crested and Raso larks. Since sexing of Raso larks is done in the field based on body size (males are larger than females), sample 26 could be an exceptionally small male or a fieldwork mistake. However, after carefully checking field notes about this individual, this seems rather unlikely. Instead, these results support a more nuanced version of the theory presented by Brooke et al. (2010). Brooke et al. (2010) suggested a fusion of homologous genetic material onto both the Z and the W chromosome, as well as suppressed recombination between the two sex chromosomes, the latter *by virtue of these chromosomes being sex chromosomes*. While general lack of recombination between sex chromosomes in birds is the norm, regions of the sex chromosomes that are similar would still be recombining (Johnson and Lachance 2012), including the

homologous pieces that fused onto the sex chromosomes as described above. Recombination could concern a large proportion of the sex chromosomes in larks, given the lack of degradation of the W chromosome frequently seen in birds (Wang et al. 2014) and the large size of both sex chromosomes in larks (Bulatova 1973). Hence, it is highly probable that after the fusion with the sex chromosomes, which took place 42.2 million years ago (Pala et al. 2012), recombination continued to happen, until it was suppressed more recently, for example through one or more inversions on the Z chromosome. If the ancestral, recombining Z still exists at low frequency in the population, it is possible, but rare, to find a homozygous female. This is the same process as for stratum formation (Wright et al. 2016), but it is not complete, either due to balancing selection, or because not enough time has elapsed yet.

The first step to investigate this hypothesis further would be to genetically sex individual 26 (we looked at read depth on the sex-linked chromosomes - where a female should have half of the read depth of a male, but because of the low coverage of the data, this was not conclusive). If sample 26 is indeed a female, it would be interesting to sequence (either RADseq or specific sequences from the scaffolds with increased heterozygosity) more Raso lark females and see how many other homozygous females can be found. This would provide more insights into the proportion of homozygous females in the population and into the types of selection at play for the maintenance of this polymorphism. New experiments could also be designed. One could genotype parents and broods to investigate the transmission of specific alleles from mothers to daughters, and from fathers to sons. Fluorescent *in situ* hybridization could be used to prove that tentative sex-linked sequences are indeed located on the sex chromosomes. Finally, it would be interesting to sequence more lark species to determine with greater precision where on the phylogenetic tree the recombination suppression event occurred. However, for the time being and for the purpose of Chapter 6 and the testing of Hypothesis A, the present study does prove that the high nucleotide diversity found in the Raso lark is in part driven by certain regions of the genome that have elevated levels of heterozygosity.

After testing Hypothesis B, it appears that, in Hypothesis A, either the estimate of 150,000 larks before the Portuguese arrival was too generous, or that the Raso lark had already undergone one or multiple bottleneck(s) before the 15th century. Hypothesis A still likely plays a role in the relatively high nucleotide diversity found in the Raso lark, but other factors are in play. The regions of the Raso lark genome that do recombine seem to have maintained the genetic diversity of a pre-contraction population

smaller than 150,000 individuals, but larger than today's.

As explained in the introduction, genetic diversity in a species can be explained not only in light of its demographic and genomic history, but also in view of its life strategy. This is of particular interest given the strong difference between the Raso lark and the skylark in this respect. Figure 6.5 is based on data from Romiguier et al. (2014), and shows the relation between neutral nucleotide diversity and maximum longevity in vertebrates. The *Alauda* larks all have higher levels of nucleotide diversity than the trend line predicts for their longevity, but they all fall well within the variation present amongst vertebrates. When compared to the other lark populations, the Raso lark exactly follows the trend predicted by the line, and has a nucleotide diversity consistent with its longevity (Figure 6.5).

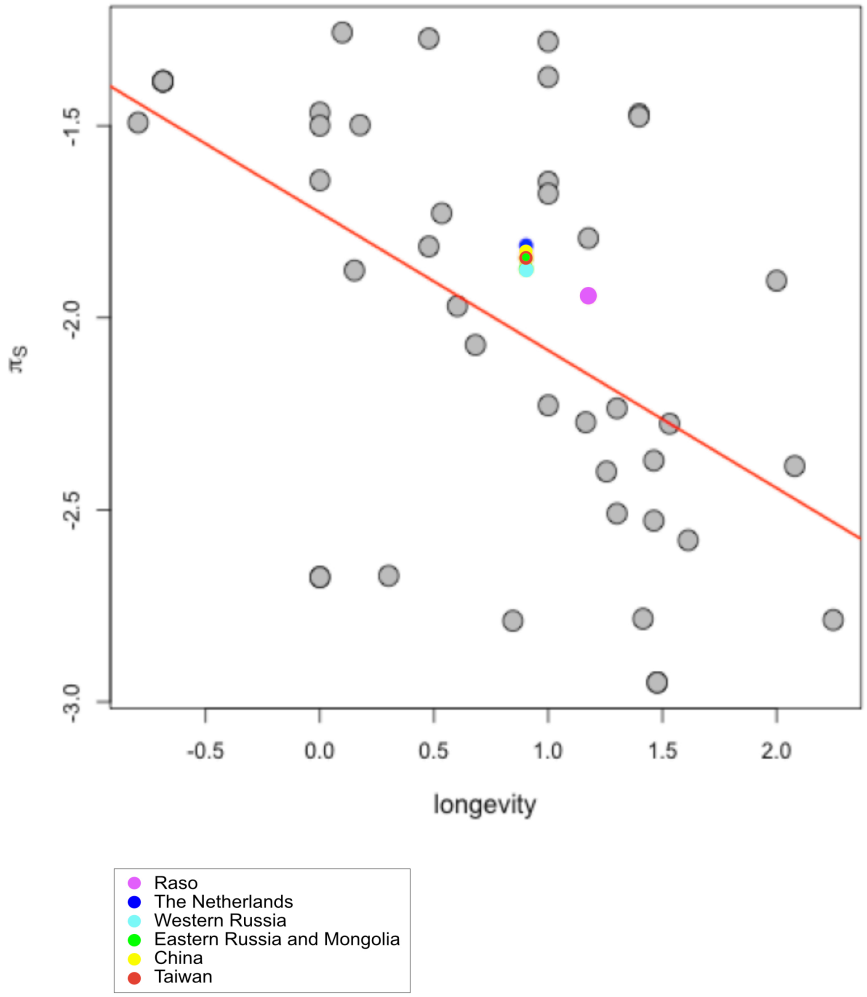


Figure 6.5 The relation between the common logarithms of neutral nucleotide diversity (π_S) and maximum longevity in vertebrates. Grey dots are based on data from Romiguier et al. (2014). The red line corresponds to a linear regression with a slope of -0.34. The coloured dots represent the relation between the common logarithms of the π

values presented in this chapter and a maximum longevity in the wild of 15 years for the Raso lark (Michael Brooke, personal communication) and 10 years for the other larks (Donald 2004).

Other genetic signatures of demographic history: high LD and kinship

Although the Raso lark population contraction and subsequent bottlenecks might not be detectable in its nucleotide diversity, it is clearly apparent in other aspects of its population genetics, including individual heterozygosity. In addition, amongst the individuals sampled - randomly and over multiple years, 81% of all pair-wise comparisons revealed genetic relatedness of third-degree or closer. In fact, almost half (46%) of the relationships were second-degree kinship, and a third were third-degree kinship. These numbers should be read in light of what we know about the recent demography of the species, with a population increasing from ≈ 20 females in 2004 to ≈ 1,500 birds in one generation (Table 1.1). Given the availability of morphological data, blood samples and recordings of long-term individual survival, the Raso lark will be a very good model species for a future study on inbreeding depression in the wild. This will require the collection of more data on individual reproductive success.

Linkage disequilibrium in the skylark and the Oriental lark (r^2 = 0.47 at its peak) is high compared to LD found in another passerine, the house finch, where the native, Eastern American population's r^2 peaks at 0.35 (Shultz et al. 2016). This might be linked to the neo sex chromosome with suppressed recombination suspected in larks (Bulatova 1973; Brooke et al. 2010; Pala et al. 2012), a mechanism that increases linkage disequilibrium. LD in the skylark and the Oriental lark is, however, lower than that found in a seabird, the black-footed albatross (r^2 = 0.60 at its peak; Dierickx et al. 2015). Linkage disequilibrium is even higher in the Raso lark, surpassing even the black-footed albatross (r^2 = 0.65 at its peak). These higher levels of LD in the Raso lark are probably the result of genetic drift due to its population contraction (Slatkin 2008). The high levels of relatedness and thus potential for inbreeding could also be a contributing factor (Slatkin 2008). The degree to which LD differs between the Raso lark and the other *Alauda* lark populations is noteworthy. While Shultz et al. (2016) found elevated genome-wide levels of LD in introduced populations of house finches, the difference between these and the native population is not more than 0.02, compared to a 0.18 difference between the Raso lark and the skylark.

While the elevated LD found by Balakrishnan and Edwards (2009) in an island population of zebra finches was accompanied by a tenfold decrease in nucleotide diversity, Shultz et al. (2016) found that a slightly elevated LD can co-exist with a more modest decrease in genetic diversity of 7-16%. This study on the Raso lark confirms this possibility, and shows that this phenomenon can be even more pronounced than previously suggested.

Loci under selection

In addition to differences in LD and genetic diversity, the Raso lark is distinct from its relative in specific locations of the genome. These differences with the skylark could be either due to the genetic effects of the population contraction and subsequent bottlenecks, or to selection in a different environment (Freeland 2005). The relative importance of these forces can be hard to disentangle. In general, selection seems to be more important; however, there is evidence that genetic drift, population contractions and bottlenecks play a larger role on islands than on the mainland (Jensen et al. 2013). Signatures of selection are likely to be masked by genetic drift if population contractions or bottlenecks happened on historical time scales (Domingues et al. 2012; Shultz et al. 2016).

BayeScan has been used before to detect selection in populations what have undergone bottlenecks (Pilot et al. 2013; Shultz et al. 2016). However, in this study, it detected only one F_{ST} outlier. In addition to the masking role played by genetic drift, this could be caused by methodological factors, such as the sparse genome sampling ability of RADseq, BayeScan's low power when few populations are compared (De Mita et al. 2013), or the fact that the reads were aligned to a skylark genome. Indeed, during alignment, the most divergent and hence most interesting regions of the Raso lark could have been discarded. The detected outlier, when mapped to the zebra finch genome, was found to be located near a gene that plays a key role in DNA repair, telomere shortening and ageing. This could be linked to two of the key differences between the Raso lark and the skylark. The former lives in a desert, with high levels of sun exposure: UV radiation is known to be one of the most common causes of DNA damage (Helleday, Eshtad and Nik-Zainal 2014). The Raso lark's lifespan is also much longer than that of the skylark. Consequently this site under selection could cautiously be linked to the picture of the Raso lark drawn in Chapters 3 and 4: that of a species investing more heavily into maintenance than its relative, the skylark.

Conclusion

This study shows that species with very small population sizes that have undergone an extreme population contraction can still present relatively high levels of nucleotide diversity, either thanks to a sufficiently large past population size and recent timing of the contraction, and/or to suppressed recombination in certain regions of the genome. However, despite this, other genetic signatures of this demographic history can be detected, including low levels of individual heterozygosity, high levels of LD and high relatedness between individuals. These results have strong implications for the conservation of the Raso lark and other species with similar demographic histories, and will be discussed in the next chapter.

Chapter 7: Conclusion

Santa Luzia Island, a second home for the Raso lark? © Paul Donald

Abstract

The Raso lark is a Critically Endangered species that is the subject of a translocation plan to the island of Santa Luzia. The research in this thesis has implications for the conservation of the species, and for the translocation in particular. Both the demographic and the genetic results help us answer some of the practical questions that need to be asked before, during, and after the translocation event. For example, when should the translocation take place? How many birds should be moved? Which birds should be chosen as founders of the new population? Finally, on a practical level, conservation measures taken for the Raso lark will cascade down to the other endangered endemic species of the Raso and Santa Luzia islands.

Current Raso lark conservation activities and plans

The Raso lark is currently found in only one location, making it extremely vulnerable to stochastic events, most notably to the introduction of invasive species such as cats or rats. These predators likely caused the extinction of the Raso lark from the other islands in Cape Verde where it used to live as well. Raso is currently the only remaining island free of invasive predators (with the possible exception of Branco Islet, 3km^2). For this reason, the IUCN classified the Raso lark as Critically Endangered on its Red List, according to criteria B1ac(iv)+2ac(iv) (IUCN 2016).

In the case of a single, potentially inbred population with no related species or subspecies with which to hybridize, Frankham, Ballou and Briscoe (2002) define four main conservation objectives. The first objective is to stop decline and increase the size of the population, the second is to maximize reproductive rate, the third is to insulate the species from environmental change, and the fourth is to establish other populations in several locations. To address the first three of these four objectives, the Desert Islands National Park - which includes the three uninhabited islands of Raso, Branco and Santa Luzia - was created in 1990. Permission from the Department of Environment is now needed to access these islands. In addition, in 1995 the Raso lark became a protected species (Donald and Brooke 2006). Sadly, the Cape Verdean government lacks the financial and human resources to effectively guard, manage and protect the species and the reserve.

To address the fourth conservation objective, in 2012 the RSPB, SPEA and Biosofera I started planning and assessing the feasibility of a reintroduction of the lark onto Santa Luzia (Geraldes et al. 2016). The habitat on Santa Luzia has already been analyzed and found suitable for the Raso lark (Geraldes et al. 2016); the partners in the project are currently working on the next step: the eradication, or at least control, of cats on what may become the lark's second home. This reintroduction will also help achieve the other conservation objectives described above. Indeed, establishing a second lark population reduces the risks linked to "having all the larks in one basket," in the case of an accidental introduction of cats or rats, for example. Moreover, by increasing the range and resources available to the Raso lark, a large increase in the total number of individuals can be expected, especially since Santa Luzia (35km^2) is much larger than Raso (7km^2). Increasing the lark's range could even increase its reproductive rate, at least temporarily, while the newly-established population expands towards the island's carrying capacity. Raso is currently very densely populated, and while this does not

seem to affect individual survival rates (see Chapter 4), it could affect reproductive output.

General IUCN reintroduction guidelines

The planning of the Raso lark translocation has been in line with the IUCN recommendations set out in a document called "Guidelines for reintroductions and other conservation translocations" (IUCN 2013; Geraldes et al. 2016). This document guides conservationists through a set of questions that need to be asked and answered during the planning of a translocation; it does not, however, aim at being a set of absolute instructions.

The IUCN guidelines start by encouraging translocation planners to prove the need for such an undertaking, including examining and eliminating all alternative conservation actions, and to set the objectives of the project. Next, the guidelines provide a framework for assessing the feasibility - biologically, socially, legally and economically - of the proposed reintroduction. In parallel, the document guides conservationists through the necessary risk assessment, including an evaluation of the risk to the source population, ecological dangers, potential associated invasions, diseases, and socio-economic hazards. Next, the document gives detailed recommendations for the practical field activities that will take place, including for the release of the founder individuals and the long-term monitoring of the new population. Finally, the IUCN guidelines emphasize the importance of disseminating information about the translocation, so that other projects worldwide can learn from both its mistakes and its successes (IUCN 2013).

In the case of the Raso lark, the project justification, feasibility assessment and risk assessment have already been performed (Geraldes et al. 2016), as well as a number of the pre-translocation fieldwork tasks (long-term monitoring of the source population, evaluating the suitability of the habitat on Santa Luzia and the on-going cat eradication).

Specific guidelines for the Raso lark based on this thesis

Pre-translocation

In order to gather the necessary fundamental biological information about the species, it is important to engage in long-term monitoring of the source population, both

before and after the translocation. In particular, the IUCN recommends collecting information about reproduction, mating systems, social behaviour, physical adaptations, individual development, parental care and population dynamics in the indigenous range (IUCN 2013). Long-term monitoring of the source population also creates a solid reference point against which to compare the new population. The Raso lark is the subject of such long-term monitoring, and a dataset covering the past 15 years has been assembled (see Chapter 2). In addition to this work, complementary research on the ecology of Santa Luzia and Raso has been undertaken by Biosfera I and SPEA, focusing mainly on the vegetation of both islands (Geraldes et al. 2016)

This thesis contributes to the health check of the source population, both demographically and genetically. Demographically, the species' high individual year-to-year survival and K-selected profile are reassuring: the lark seems very well adapted to the arid conditions on Raso, can withstand numerous dry years in a row (see Chapters 3 and 4), and its population size can increase rapidly as soon as there is enough rainfall (Brooke et al. 2012; Figure 4.1). Genetically, the main question to ask is: has the species lost too much genetic diversity? While its nucleotide diversity is currently still high, this should not be considered "good news" for the Raso lark. Because the population contraction of the species was relatively recent, its effects on π are not fully apparent yet, and nucleotide diversity is expected to continue dropping for many generations. Furthermore, while some individuals have increased heterozygosity in certain parts of their genome, many do not. Even in the case of the more heterozygous birds, this only concerns 16% of their genome - the remaining 84% is much less heterozygous than in other *Alauda*. The population contraction has left additional signatures in the Raso lark's genome, such as very high levels of linkage disequilibrium. The RADseq data also showed that individuals on Raso are highly related, warning us of potential inbreeding depression in the species (see Chapter 6).

Capture and release

In addition to this health check of the source population, this thesis helps us answer some crucial practical questions for the design of the translocation:

Should the translocation happen sooner, or later? There are three strong arguments in favour of a translocation in the near future. First, as time goes by, the likelihood of a catastrophic event or a predator invasion increases. Second, it is best to

draw individuals from the source population while it is still large. Population size estimates show that its peak - 1558 birds in 2011 - has already been missed, and that it is now slowly declining (Table 1.1). Third, when a species has undergone a bottleneck, it is best to increase the population size again as quickly as possible, because the shorter these events, the smaller their impact on genetic diversity (Frankham, Ballou and Briscoe 2002). While this might already be too late for the Raso lark - the main population contraction happened in the 15th century, and the genome carries clear signatures of the event - it is still worth trying to minimize negative consequences, since nucleotide diversity is still high.

However, there is a major obstacle to an immediate translocation: cat extermination on Santa Luzia is in progress, but not yet fully completed. Full removal of invasive species is well-known to be a very lengthy process: while initial progress is rapid, it then slows, and it can take many years before the last few individuals are found and removed (Michael Brooke, personal communication). The danger of exposing some Raso larks to a few predators needs to be balanced against the danger of "leaving all the birds in the same basket" on Raso and against the missed opportunity of a historically large source population. According to Annex 3.10 of the IUCN guidelines (2013), "a trial release may answer uncertainties (…), but should only be contemplated where all formal requirements have been met, where the consequences will be suitably monitored and will be used to refine further release design, and any unacceptable impacts can be mitigated or reversed." With a controlled cat population on Santa Luzia and a suitable monitoring scheme in place, this could be the answer to the dilemma.

At what time of the year should the translocation take place? First and foremost, the procedure should take place outside the breeding season, so as not to disturb reproduction and minimize the birds' desire to fly back from Santa Luzia to Raso, since the two islands are only 18km apart and within eyesight of each other. The timing of the breeding season can be difficult to predict in this species, so Raso should be checked for breeding activity before individuals are removed. However, breeding is most likely to happen between September and December, which corresponds the Cape Verdean rainy season. Ideally, the translocation would be followed by a season of abundant rainfall which, despite potentially slightly decreasing adult survival, would prompt a burst of reproduction. However, given the high inter-annual variation in rainfall in Cape Verde, it is impossible to plan for this. While adults have high survival rates (> 80%) and the new population could persist for multiple years without breeding, a population of 50

individuals would nevertheless dwindle to just 17 in five years. Hence, the new population will have to be monitored closely, and might need reinforcing. Providing water to the birds in a few locations on Santa Luzia, in the hope of prompting breeding, might be another option. However, if that option is chosen, potential risks - for example, increased predation at water stations - will have to be evaluated and mitigated.

How many translocation events will be needed? As explained above, this will in part depend on how quickly the birds start breeding on Santa Luzia. Additionally, the Santa Luzia population could be reinforced opportunistically when the Raso population is large. Releasing individuals over several years may help surmount problems linked to inter-annual variation in rainfall and the occurrence of infrequent but severe natural catastrophes (IUCN 2013). Since the Santa Luzia population will initially be smaller than the Raso population, it will be more vulnerable to stochastic events. Furthermore, the Santa Luzia population will also be more at risk from a genetic perspective. Bottlenecks prevent founders from contributing equally to the new population's gene pool. Releasing additional birds would limit the genetic risk of such bottlenecks, as well as reduce differentiation over time between the two islands. However, while some authors do suggest that multiple introductions reduce the founder effect (Dlugosch and Parker 2008), Clegg et al. (2002) argue that, to reduce founder effects, it is better to reintroduce many individuals the first time rather than translocate a few birds many times over, based on their study of natural colonization events by the silvereye *Zosterops lateralis*.

How many individuals should be translocated? The answer to this question depends in part on the considerations laid out above. It also depends on the current levels of genetic diversity in the source population. When I started researching this topic, the favoured hypothesis was that genetic diversity in the Raso lark was low and that very few birds would need to be translocated in order to capture virtually all of the genetic diversity of the source population. For example, if the calculations in Chapter 6 had yielded Ne = 40 (the number of individuals to which the population has apparently been reduced multiple times in the past; Table 1.1), translocating 50 individuals would have been more than sufficient from a genetic perspective. However, my calculations generated estimates for Ne four orders of magnitude larger than this. Since much of this genetic diversity is still present in the population - due to deep coalescence, incomplete lineage sorting and suppressed recombination - increasing the number of founders

would increase the chances of the different lineages being represented in the new population. This would, in turn, reduce the chances of inbreeding depression and inability to adapt to environmental change. Since the Raso lark is probably already at risk of this due to the population contraction that it endured, it is important not to endanger it even further by creating a bottleneck for the new population. Given the two islands' proximity, one can hope for gene flow between the two populations, which would reduce that risk. All in all, the most important consideration for the decision on how many individuals to translocate is the trade-off between maximizing the chance of establishment of the new population and minimizing harm to the source population.

Which individuals should be translocated? Given the proximity of Raso to Santa Luzia, newly translocated birds could fly back to Raso. Two ways to mitigate this risk are to translocate birds outside of the breeding season and to target young birds that do not hold territories on Raso yet. Choosing young individuals has the added advantage of maximizing their years living (and hence breeding) on Santa Luzia. Practically speaking, the sample of birds can be shifted towards younger birds by examining claw damage, but some older birds could still be sampled this way. Choosing birds based on their genotype is not feasible in practice; however, blood samples should be taken for later genotyping. While results in Chapter 4 hinted at the fact that smaller birds have higher survival rates than larger birds, choosing smaller birds for the reintroduction could mean favouring certain genotypes over others, and hence limit the evolutionary adaptability of the new population. Consequently, it is better in this case to sample the source population randomly. Finally, I recommend an even sex ratio for the founder population, because, although Chapter 3 found that year-to-year survival depended on sex, it also showed that this interaction and its direction was time-dependent. If after a few years the sex ratio on Santa Luzia becomes skewed (as has happened on Raso in the past), the sex ratio of the reinforcement cohorts can be adjusted accordingly.

Where on Santa Luzia should the founders be released? The IUCN guidelines (2013) recommend releasing the founders at multiple sites, in the hope that this will increase the chances of selecting a favourable habitat and help avoid localized disturbance events.

Post-translocation

The work done about the Raso lark before and during this thesis should continue, both on Raso and on Santa Luzia. Birds on both islands should be colour-ringed, to help researchers estimate population sizes, demographic patterns, individual survival rates and inbreeding depression. Taking and storing blood samples from the founders is essential. In addition, ecological monitoring on Santa Luzia, as well as a genetic study to calculate the number of founders contributing to subsequent generations and the extent to which individuals from reinforcements are supplying genes to the new population, should be undertaken. Together, this covers the post-release monitoring themes recommended by the IUCN (2013).

Other considerations

Conservation measures other than the translocation, such as cat and rodent eradication, ecosystem monitoring or the enforcement of the natural reserve regulations, are of utmost importance for the species, and should be a priority. Every effort should be made to prevent invasive species colonising Raso. Currently, very few people access the island, apart from fishermen and the NGO Biosfera I which, every year, stays on Raso for a few months to protect the endemic, Near Threatened Cape Verde shearwater *Calonectris edwardsii* from overharvesting. While it is unlikely that cats could be imported accidentally this way, mice and rats could find their way onto Raso *via* the little boats that people use. On the other hand, in the absence of any visitor control by the Cape Verdean government, the presence of Biosfera I helps monitor the situation on Raso.

Many of the theoretical considerations above are applicable to other reintroduction projects around the world. On a practical level, conservation measures for the Raso lark will also benefit other local species, such as the Cape Verde shearwater mentioned above, the Near Threatened Raso wall gecko *Tarentola raziana* endemic to the Desert Islands, or the Endangered giant wall gecko *Tarentola gigas* endemic to Raso and Branco (Geraldes et al. 2016; Geraldes and Melo 2016; IUCN 2016). Yet another Cape Verde endemic, the Cape Verde giant skink *Macroscincus coctei*, is a stark warning of what could happen if the Desert Islands are not protected. Like the Raso lark, this species lived on numerous islands in Cape Verde until the arrival of the Portuguese and their invasive species. At that point, its range was reduced

to the three Desert Islands, where it survived until the 19th century. It was last seen by Boyd Alexander on Raso, in 1897, the year that he discovered the Raso lark (Andreone and Gavetti 1998). A century later he would doubtless be horrified to discover that the skink is no more, and anxious to ensure that the bird species he discovered does not follow the reptile into oblivion. My hope is that this thesis will help provide some of the tools needed to quell Boyd Alexander's anxiety.

The Extinct Cape Verde giant skink (Tremeau de Rochebrune 1883).

References

Akaike, H. 1973. “Information Theory and an Extension of the Maximum Likelihood Principle.” *Proceeding of the Second International Symposium on Information Theory* 1: 267-81.

Alexander, B. 1898. “An Ornithological Expedition to the Cape Verde Islands.” *Ibis* 40: 74-118.

Alström, P, KN Barnes, U Olsson, FK Barker, P Bloomer, AA Khan, MA Qureshi, A Guillaumet, P-A Crochet, and PG Ryan. 2013. “Multilocus Phylogeny of the Avian Family Alaudidae (Larks) Reveals Complex Morphological Evolution, Non-Monophyletic Genera and Hidden Species Diversity.” *Molecular Phylogenetics and Evolution* 69: 1043-56.

Altschul, SF, W Gish, W Miller, EW Myers, and DJ Lipman. 1990. “Basic Local Alignment Search Tool.” *Journal of Molecular Biology* 215: 403-10.

Amstrup, S, TL McDonald, and BFJ Manly. 2005. *Handbook of Capture-Recapture Analysis*. Woodstock: Princeton University Press.

Anderson, DR, and KP Burnham. 1999. “General Strategies for the Analysis of Ringing Data.” *Bird Study* 46: S261-70.

Andreone, F, and E Gavetti. 1998. “Some Remarkable Specimens of the Giant Cape Verde Skink, *Macroscincus Coctei* (Duméril & Bibron, 1839), with Notes about Its Distribution and Causes of Its Possible Extinction.” *Italian Journal of Zoology* 65: 413-21.

Andrews, KR, JM Good, MR Miller, G Luikart, and PA Hohenlohe. 2016. “Harnessing the Power of RADseq for Ecological and Evolutionary Genomics.” *Nature Reviews Genetics* 17: 81-92.

Andrews, S. 2010. “FastQC: A Quality Control Tool for High Throughput Sequence Data.” http://www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/.

Balakrishnan, CN, and SV Edwards. 2009. “Nucleotide Variation, Linkage Disequilibrium and Founder-Facilitated Speciation in Wild Populations of the Zebra Finch (*Taeniopygia Guttata*).” *Genetics* 181: 645-60.

Barrowclough, GF, J Cracraft, J Klicka, and RM Zink. 2016. “How Many Kinds of Birds Are There and Why Does It Matter?” *PLoS ONE* 11: e0166307.

Bellard, C, C Bertelsmeier, P Leadley, W Thuiller, and F Courchamp. 2012. “Impacts of Climate Change on the Future of Biodiversity.” *Ecology Letters* 15: 365-77.

BirdLife International. 2011. “We Have Lost over 150 Bird Species since 1500.” http://datazone.birdlife.org/sowb/casestudy/we-have-lost-over-150-bird-species-since-1500.

BirdLife International. 2016. “Species Factsheet.” http://www.birdlife.org/datazone/species- factsheet.php?id=3957.

Blake, JG, and BA Loiselle. 2008. “Estimates of Apparent Survival Rates for Forest Birds in Eastern Ecuador.” *Biotropica* 40: 485-93.

Blake, JG, and BA Loiselle. 2013. “Apparent Survival Rates of Forest Birds in Eastern Ecuador Revisited: Improvement in Precision but No Change in Estimates.” *PLoS ONE* 8: e81028

Bolger, AM, M Lohse, and B Usadel. 2014. “Trimmomatic: A Flexible Trimmer for Illumina Sequence Data.” *Bioinformatics* 30: 2114-20.

Brekke, P, PM Bennett, J Wang, N Pettorelli, and JG Ewen. 2010. “Sensitive Males: Inbreeding Depression in an Endangered Bird.” *Proceedings of the Royal Society B: Biological Sciences* 277: 3677-84.

Brito, PH, and SV Edwards. 2009. “Multilocus Phylogeography and Phylogenetics Using Sequence-Based Markers.” *Genetica* 138: 465.

Brook, BW, DW Tonkyn, JJ O’Grady, and R Frankham. 2002. “Contribution of Inbreeding to the Extinction Risk in Threatened Species.” *Conservation Ecology* 6: 16.

Brooke, M de L, and M Bolton. 2002. “Raso Lark Studies 2002.” University of Cambridge. Unpublished report.

Brooke, M de L, and EG Dierickx. 2013. “Raso Lark Studies 2013.” University of Cambridge. Unpublished report.

Brooke, M de L, and EG Dierickx. 2014. “Raso Lark Studies 2014.” University of Cambridge. Unpublished report.

Brooke, M de L, and TP Flower. 2011. “Raso Lark Studies 2011.” University of Cambridge. Unpublished report.

Brooke, M de L, and LV Gregory. 2016. “Raso Lark Studies 2016.” University of Cambridge. Unpublished report.

Brooke, M de L, and S Hille. 2003. “Raso Lark Studies 2003.” University of Cambridge. Unpublished report.

Brooke, M de L, and MC Mainwaring. 2015. “Raso Lark Studies 2015.” University of Cambridge. Unpublished report.

Brooke, M de L, and JA Welbergen. 2005. “Raso Lark Studies 2005.” University of Cambridge. Unpublished report.

Brooke, M de L, and JA Welbergen. 2010. “Raso Lark Studies 2010.” University of Cambridge. Unpublished report.

Brooke, M de L, JA Welbergen, MC Mainwaring, M Van Der Velde, AMF Harts, J Komdeur, and W Amos. 2010. “Widespread Translocation from Autosomes to Sex Chromosomes Preserves Genetic Variability in an Endangered Lark.” *Journal of Molecular Evolution* 70: 242-6.

Brooke, M de L, TP Flower, EM Campbell, MC Mainwaring, S Davies, and JA Welbergen. 2012. “Rainfall-Related Population Growth and Adult Sex Ratio Change in the Critically Endangered Raso Lark (*Alauda Razae*).” *Animal Conservation* 15: 466-71.

Brooks, DJ, S Cramp, E Dunn, R Gillmor, J Hall-Craggs, B Hillcoat, PAD Hollom, EM Nicholson, CM Perrins, CS Roselaar, et al. 1988. *Handbook of the Birds of Europe, the Middle East and North Africa*. Oxford: Oxford University Press.

Bulatova, NS. 1973. “Unusually Large Sex Chromosomes in Some Larks (Aves: Alaudidae).” *Mammalian Chromosomes Newsletter* 14: 150-1.

Burnham, KP, and DR Anderson. 2004. “Multimodel Inference: Understanding AIC and BIC in Model Selection.” *Sociological Methods and Research* 33: 261-304.

Burnham, KP, and DR Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer.

Butchart, SHM, AJ Stattersfield, and TM Brooks. 2006. “Going or Gone: Defining ‘Possibly Extinct’ Species to Give a Truer Picture of Recent Extinctions.” *Bulletin of the British Ornithological Club* 126: 7-24.

Cade, BS. 2015. “Model Averaging and Muddled Multimodel Inferences.” *Ecological Society of America* 96: 2370-82.

Campbell, B, and E Lack. 1985. *A Dictionary of Birds*. London: T & AD Poyser.

Campbell, RLW, LM Van Damme, and SR Johnson. 1997. *The Birds of North America*. Ithaca: Cornell Lab of Ornithology.

Canal, D, D Serrano, and J Potti. 2014. “Exploring Heterozygosity-Survival Correlations in a Wild Songbird Population: Contrasting Effects between Juvenile and Adult Stages.” *PLoS ONE* 9: e105020.

Carthey, AJR, and PB Banks. 2014. “Naïveté in Novel Ecological Interactions: Lessons from Theory and Experimental Evidence.” *Biological Reviews* 89: 932-49.

Catchen, JM, A Amores, P Hohenlohe, W Cresko, JH Postlethwait, and D-J De Koning. 2011. "Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences." *G3: Genes, Genomes, Genetics* 1: 171-82.

Catchen, J, PA Hohenlohe, S Bassham, A Amores, and WA Cresko. 2013. "Stacks: An Analysis Tool Set for Population Genomics." *Molecular Ecology* 22: 3124-40.

Catry, P, RA Phillips, B Phalan, and JP Croxall. 2006. "Senescence Effects in an Extremely Long-Lived Bird: The Grey-Headed Albatross *Thalassarche Chrysostoma." Proceedings of the Royal Society B: Biological Sciences* 273: 1625-30.

Chaverri, G, and TH Kunz. 2011. "All-Offspring Natal Philopatry in a Neotropical Bat." *Animal Behaviour* 82: 1127-33.

Christiansen, DG, C Jakob, M Arioli, S Roethlisberger, and H-U Reyer. 2010. "Coexistence of Diploid and Triploid Hybrid Water Frogs: Population Differences Persist in the Apparent Absence of Differential Survival." *BMC Ecology* 10: 14.

CIA. 2016. "The World Factbook: Cabo Verde." https://www.cia.gov/library/publications/the-world-factbook/geos/cv.html.

Clegg, SM, SM Degnan, J Kikkawa, C Moritz, A Estoup, and IPF Owens. 2002. "Genetic Consequences of Sequential Founder Events by an Island-Colonizing Bird." *Proceedings of the National Academy of Sciences of the United States of America* 99: 8127-32.

Cohas, A, C Bonenfant, B Kempenaers, and D Allainé. 2009. "Age-Specific Effect of Heterozygosity on Survival in Alpine Marmots, *Marmota Marmota*." *Molecular Ecology* 18: 1491-1503.

Cooch, E, and G White. 2006. *Program MARK: A Gentle Introduction*. http://www.phidot.org/software/mark/docs/book/

Cooper, WE, RA Pyron, and T Garland. 2014. "Island Tameness: Living on Islands Reduces Flight Initiation Distance." *Proceedings of the Royal Society B: Biological Sciences* 281: 20133019.

Coues, E. 1882. *A Check List of North American Birds*. Boston: Estes and Lauriat.

Covas, R. 2012. "Evolution of Reproductive Life Histories in Island Birds Worldwide." *Proceedings of the Royal Society B: Biological Sciences* 279: 1531-7.

Cox, WA, FR Thompson, AS Cox, and J Faaborg. 2014. "Post-Fledging Survival in Passerine Birds and the Value of Post-Fledging Studies to Conservation." *Journal of Wildlife Management* 78: 183-93.

Darwin, C. 1839. *Narrative of the Surveying Voyages of His Majesty's Ships Adventure and Beagle between the Years 1826 and 1836, Describing Their Examination of the Southern Shores of South America, and the Beagle's Circumnavigation of the Globe. Journal and Remarks*. London: Henry Colburn.

Davey, JL, and MW Blaxter. 2010. "RADseq: Next-Generation Population Genetics." *Briefings in Functional Genomics* 9: 416-23.

De Mita, S, AC Thuillet, L Gay, N Ahmadi, S Manel, J Ronfort, and Y Vigouroux. 2013. "Detecting Selection along Environmental Gradients: Analysis of Eight Methods and Their Effectiveness for Outbreeding and Selfing Populations." *Molecular Ecology* 22: 1383-99.

del Hoyo, J, and NJ Collar. 2014. *HBW and BirdLife International Illustrated Checklist of the Birds of the World. Volume 1: Non-Passerines*. Barcelona: Lynx Edicions.

Descamps, S, HG Gilchrist, J Bêty, EI Buttler, and MR Forbes. 2009. "Costs of Reproduction in a Long-Lived Bird: Large Clutch Size Is Associated with Low Survival in the Presence of a Highly Virulent Disease." *Biology Letters* 5: 278-81.

Dierickx, EG, AJ Shultz, F Sato, T Hiraoka, and SV Edwards. 2015. "Morphological and Genomic Comparisons of Hawaiian and Japanese Black-Footed Albatrosses (*Phoebastria Nigripes*) Using Double Digest RADseq: Implications for Conservation." *Evolutionary Applications* 8: 662-78.

Dlugosch, KM, and IM Parker. 2008. "Founding Events in Species Invasions: Genetic Variation, Adaptive Evolution, and the Role of Multiple Introductions." *Molecular Ecology* 17: 431-49.

Doherty, TS, AS Glen, DG Nimmo, EG Ritchie, and CR Dickman. 2016. "Invasive Predators and Global Biodiversity Loss." *Proceedings of the National Academy of Sciences of the United States of America* 113: 11261-5.

Doi, H, K Yoshida, T Yasuda, M Fukuda, Y Fukuda, H Morita, SI Ikeda, R Kato, Y Tsurusaki, N Miyake, et al. 2011. "Exome Sequencing Reveals a Homozygous SYT14 Mutation in Adult-Onset, Autosomal-Recessive Spinocerebellar Ataxia with Psychomotor Retardation." *American Journal of Human Genetics* 89: 320-7.

Domingues, VS, YP Poh, BK Peterson, PS Pennings, JD Jensen, and HE Hoekstra. 2012. "Evidence of Adaptation from Ancestral Variation in Young Populations of Beach Mice." *Evolution* 66: 3209-23.

Donald, PF. 2004. *The Skylark*. London: T & AD Poyser.

Donald, PF. 2007. "Adult Sex Ratios in Wild Bird Populations." *Ibis* 149: 671-92.

Donald, PF, and M de L Brooke. 2006. “An Unlikely Survivor: The Peculiar Natural History of the Raso Lark.” *British Birds* 999: 420-30.

Donald, PF, M De Ponte, MJ Pitta Groz, and R Taylor. 2003. “Status, Ecology, Behaviour and Conservation of Raso Lark *Alauda Razae*.” *Bird Conservation International* 13: 13-28.

Donald, PF, S Hille, M de L Brooke, R Taylor, CE Wells, M Bolton, and T Marlow. 2007. “Sexual Dimorphism, Niche Partitioning and Social Dominance in the Feeding Ecology of the Critically Endangered Raso Lark *Alauda Razae*.” *Ibis* 149: 848-52.

Earl, DA, and BM vonHoldt. 2012. “STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method.” *Conservation Genetics Resources* 4: 359-61.

Edmands, S. 2007. “Between a Rock and a Hard Place: Evaluating the Relative Risks of Inbreeding and Outbreeding for Conservation and Management.” *Molecular Ecology* 16: 463-75.

Edwards, SV, and S Bensch. 2009. “Looking Forwards or Looking Backwards in Avian Phylogeography? A Comment on Zink and Barrowclough 2008.” *Molecular Ecology* 18: 2930-3.

Ellegren, H. 2007. “Molecular Evolutionary Genomics of Birds.” *Cytogenetic and Genome Research* 117: 120-30.

Espejel, S, M Martín, P Klatt, J Martín-Caballero, JM Flores, and MA Blasco. 2004. “Shorter Telomeres, Accelerated Ageing and Increased Lymphoma in DNA-PKcs-Deficient Mice.” *EMBO Reports* 5: 503-9.

Evanno, G, S Regnaut, and J Goudet. 2005. “Detecting the Number of Clusters of Individuals Using the Software STRUCTURE: A Simulation Study.” *Molecular Ecology* 14: 2611-20.

Ferreira, MT, P Cardoso, PAV Borges, R Gabriel, EB de Azevedo, F Reis, MB Araújo, and RB Elias. 2016. “Effects of Climate Change on the Distribution of Indigenous Species in Oceanic Islands (Azores).” *Climate Change* 138: 603-15.

Flegg, J, and D Hosking. 2002. *Birds of Britain and Europe*. Wiltshire: Selectabook Ltd.

Flicek, P, MR Amode, D Barrell, K Beal, S Brent, D Carvalho-Silva, P Clapham, G Coates, S Fairley, S Fitzgerald, et al. 2012. “Ensembl 2012.” *Nucleic Acids Research* 40: D84-90.

Foll, M, and O Gaggiotti. 2008. “A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective.” *Genetics* 180: 977-93.

Frankham, R. 1998. “Inbreeding and Extinction: Island Populations.” *Conservation Biology* 12: 665-75.

Frankham, R, JD Ballou, and DA Briscoe. 2002. *Introduction to Conservation Genetics*. Cambridge: Cambridge University Press.

Freeland, JR. 2005. *Molecular Ecology*. Chichester: John Wiley & Sons.

Geraldes, P, J Kelly, T Melo, and PV Donald. 2016. “The Restoration of Santa Luzia, Republic of Cabo Verde, Feasibility Study and Restoration Action Plan 2016-2020.” Sociedade Portuguesa para o Estodo das Aves.

Geraldes, P, and T Melo. 2016. “The Restoration of Santa Luzia, Republic of Cabo Verde, Terrestrial Reptile Monitoring Report 2013-2015.” Sociedade Portuguesa para o Estodo das Aves.

Gnerre, S, I Maccallum, D Przybylski, FJ Ribeiro, JN Burton, BJ Walker, T Sharpe, G Hall, TP Shea, S Sykes, et al. 2011. “High-Quality Draft Assemblies of Mammalian Genomes from Massively Parallel Sequence Data.” *Proceedings of the National Academy of Sciences of the United States of America* 108: 1513-8.

Gossmann, TI, PD Keightley, and A Eyre-Walker. 2012. “The Effect of Variation in the Effective Population Size on the Rate of Adaptive Molecular Evolution in Eukaryotes.” *Genome Biology and Evolution* 4: 658-67.

Hagman, M, BL Phillips, and R Shine. 2009. “Fatal Attraction: Adaptations to Prey on Native Frogs Imperil Snakes after Invasion of Toxic Toads.” *Proceedings of the Royal Society B: Biological Sciences* 276: 2813-8.

Hanssen, SA, D Hasselquist, I Folstad, and KE Erikstad. 2005. “Cost of Reproduction in a Long-Lived Bird: Incubation Effort Reduces Immune Function and Future Reproduction.” *Proceedings of the Royal Society B: Biological Sciences* 272: 1039-46.

Harrison, CJO. 1966. “The Validity of Some Genera of Larks (Alaudidae).” *Ibis* 108: 573-83.

Hartert, E. 1922. *Die Vögel Der Paläarktischen Fauna: Systematische Übersicht Der in Europa, Nord-Asien Und Der Mittelmeerregion Vorkommenden Vögel*. Berlin: Friedländer & Sohn.

Hazevoet, CJ. 1995. *The Birds of the Cape Verde Islands: An Annotated Checklist*. Peterborough: British Ornithologist’s Union.

Hazevoet, CJ. 2012. “Seventh Report on Birds from the Cape Verde Islands, Including Records of Nine Taxa New to the Archipelago.” *Zoologia Caboverdiana* 3: 1-28.

Hazevoet, CJ. 2014. “Eighth Report on Birds from the Cape Verde Islands, Including Records of Nine Taxa New to the Archipelago.” *Zoologia Caboverdiana* 5: 29-56.

Hedrick, PW. 2007. “Balancing Selection.” *Current Biology* 17: R230-1.

Helleday, T, S Eshtad, and S Nik-Zainal. 2014. “Mechanisms Underlying Mutational Signatures in Human Cancers.” *Nature Reviews Genetics* 15: 585-98.

Ho, S-R, CS Mahanic, Y-J Lee, and W-C Lin. 2014. “RNF144A, an E3 Ubiquitin Ligase for DNA-PKcs, Promotes Apoptosis during DNA Damage.” *Proceedings of the National Academy of Sciences of the United States of America* 111: E2646-55.

Holmes, DJ, R Flückiger, and SN Austad. 2001. “Comparative Biology of Aging in Birds: An Update.” *Experimental Gerontology* 36:869-83.

Horrocks, NPC, A Hegemann, KD Matson, K Hine, S Jaquier, M Shobrak, JB Williams, JM Tinbergen, and BI Tieleman. 2012. “Immune Indexes of Larks from Desert and Temperate Regions Show Weak Associations with Life History but Stronger Links to Environmental Variation in Microbial Abundance.” *Physiological and Biochemical Zoology* 85: 504-15.

Hudson, R. 2002. “ms - a Program for Generating Samples Under Neutral Models.” *Bioinformatics* 2002: 337-8.

Hughes, AR, BD Inouye, MTJ Johnson, N Underwood, and M Vellend. 2008. “Ecological Consequences of Genetic Diversity.” *Ecology Letters* 11: 609-23.

Huson, DH, and D Bryant. 2006. “Application of Phylogenetic Networks in Evolutionary Studies.” *Molecular Biology and Evolution* 24: 254-67.

IUCN. 2013. “Guidelines for Reintroductions and Other Conservation Translocations.” https://www.iucn.org/content/new-guidelines-conservation-translocations-published-iucn.

IUCN. 2017. “IUCN Red List of Threatened Species. Version 2016.3.” http://www.redlist.org/.

Jakobsson, M, and NA Rosenberg. 2007. “CLUMPP: A Cluster Matching and Permutation Program for Dealing with Label Switching and Multimodality in Analysis of Population Structure.” *Bioinformatics* 23: 1801-6.

Jensen, H, R Moe, IJ Hagen, AM Holand, J Kekkonen, J Tufto, and BE Sæther. 2013. “Genetic Variation and Structure of House Sparrow Populations: Is There an Island Effect?” *Molecular Ecology* 22: 1792-1805.

Johnson, TH, and AJ Stattersfield. 1990. “A Global Review of Island Endemic Birds.” *Ibis* 132: 167-80.

Johnson, WE, DP Onorato, ME Roelke, ED Land, M Cunningham, RC Belden, R McBride, D Jansen, M Lotz, D Shindle, et al. 2010. “Genetic Restoration of the Florida Panther.” *Science* 329: 1641-5.

Johnson, NA, and J Lachance. 2013. “The Genetics of Sex Chromosomes: Evolution and Implications for Hybrid Incompability.” 2012. *Annals of the New York Academy of Sciences 1256*: E1-22.

Johnston, JP, WJ Peach, RD Gregory, and SA White. 1997. “Survival Rates of Tropical and Temperate Passerines: A Trinidadian Perspective.” *The American Naturalist* 150: 771-89.

Jombart, T, and I Ahmed. 2011. “Adegenet 1.3-1: New Tools for the Analysis of Genome-Wide SNP Data.” *Bioinformatics* 27: 3070-1.

Jones, OR, JM Gaillard, S Tuljapurkar, JS Alho, KB Armitage, PH Becker, P Bize, J Brommer, A Charmantier, M Charpentier, et al. 2008. “Senescence Rates Are Determined by Ranking on the Fast-Slow Life-History Continuum.” *Ecology Letters* 11: 664-73.

Kamilar, J, and A Pokempner. 2008. “Does Body Mass Dimorphism Increase Male-female Dietary Niche Separation? A Comparative Study of Primates.” *Behaviour* 145: 1211-34.

Keller, LF, and DM Waller. 2002. “Inbreeding Effects in Wild Populations.” *Trends in Ecology and Evolution* 17: 230-41.

Langmead, B, and SL Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9: 357-9.

Lebreton, JD, KP Burnham, J Clobert, and DR Anderson. 1992. “Modeling Survival and Testing Biological Hypotheses Using Marked Animals: A Unified Approach with Case Studies.” *Ecological Monographs* 62: 67-118.

Lee, WG, JR Wood, and GM Rogers. 2010. “Legacy of Avian-Dominated Plant-Herbivore Systems in New Zealand.” *New Zealand Journal of Ecology* 34: 28-47.

Leech, TJ, E Craig, B Beaven, DK Mitchell, and PJ Seddon. 2007. “Reintroduction of Rifleman *Acanthisitta Chloris* to Ulva Island, New Zealand: Evaluation of Techniques and Population Persistence.” *Oryx* 41: 369-75.

Lowe, S, M Browne, S Boudjelas, and M De Poorter. 2000. “100 of the World’s Worst Invasive Alien Species: A Selection from the Global Invasive Species Database.” *Aliens* 12: S1-11.

Manne, LL, TM Brooks, and SL Pimm. 1999. "Relative Risk of Extinction of Passerine Birds on Continents and Islands." *Nature* 399: 258-61.

Mateo, JA, LF Lópex Jurado, and P Geniez. 2009. "Historical Distribution of Razo Lark *Alauda Razae* in the Cape Verde Archipelago." *Alauda* 77: 309-12.

McCreless, EE, DD Huff, DA Croll, BR Tershy, ND Holmes, DR Spatz, SHM Butchart, and C Wilcox. 2016. "Past and Predicted Future Impact of Invasive Alien Mammals on Island Threatened Vertebrate Populations." *Nature Communications* 7: 1-11.

Meinertzhagen, R. 1951. "Review of the Alaudidae." *Journal of Zoology* 121: 81-132.

Merrill, R. 2014. "RAD Library Preparation Pstl Protocol Based on Genepool v1.2." Cambridge. Unpublished protocol.

Ministry of Environment Housing and Territory Planning of Cape Verde. 2011. "Climate Change Vulnerability Assessment of Cape Verde: Summary for Policy Makers." Praia.

Monticelli, D, R Ceia, R Heleno, H Laborda, S Timóteo, D Jareño, GM Hilton, and JA Ramos. 2010. "High Survival Rate of a Critically Endangered Species, the Azores Bullfinch *Pyrrhula Murina*, as a Contribution to Population Recovery." *Journal of Ornithology* 151: 627-36.

Mounce, HL, KJ Iknayan, DL Leonard, KJ Swinnerton, and JJ Groombridge. 2014. "Management Implications Derived from Long Term Re-Sight Data: Annual Survival of the Maui Parrotbill *Pseudonestor Xanthophrys*." *Bird Conservation International* 24: 316-26.

Nadeau, NJ, M Ruiz, P Salazar, B Counterman, JA Medina, H Ortiz-Zuazaga, A Morrison, WO McMillan, CD Jiggins, and R Papa. 2014. "Population Genomics of Parallel Hybrid Zones in the Mimetic Butterflies, *H. Melpomene* and *H. Erato*." *Genome Research* 24: 1316-33.

Nair, P. 2014. "Conservation Genomics." *Proceedings of the National Academy of Sciences of the United States of America* 111: 569.

Newton, I. 1998. *Population Limitation in Birds*. San Diego: Academic Press.

O'Grady, JJ, BW Brook, DH Reed, JD Ballou, DW Tonkyn, and R Frankham. 2006. "Realistic Levels of Inbreeding Depression Strongly Affect Extinction Risk in Wild Populations." *Biological Conservation* 133: 42-51.

Pala, I, S Naurin, M Stervander, D Hasselquist, S Bensch, and B Hansson. 2012. "Evidence of a Neo-Sex Chromosome in Birds." *Heredity* 108: 264-72.

Parmesan, C. 2006. “Ecological and Evolutionary Responses to Recent Climate Change.” *Annual Review of Ecology Evolution and Systematics* 37: 637-69.

Péron, G, O Gimenez, A Charmantier, J-M Gaillard, and P-A Crochet. 2010. “Age at the Onset of Senescence in Birds and Mammals Is Predicted by Early-Life Performance.” *Proceedings of the Royal Society B: Biological Sciences* 277: 2849-56.

Pickrell, JK, and JK Pritchard. 2012. “Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data.” *PLoS Genetics* 8: 1-17.

Pilot, M, C Greco, BM vonHoldt, B Jędrzejewska, E Randi, W Jędrzejewski, VE Sidorovich, E Ostrander, and RK Wayne. 2014. “Genome-Wide Signatures of Population Bottlenecks and Diversifying Selection in European Wolves.” *Heredity* 112: 428-42.

Pim, J. 2008. “Crustal Structure and Origin of the Cape Verde Rise.” *Earth and Planetary Science Letters* 272: 422-8.

Prevosti, A, J Ocaña, and G Alonso. 1975. “Distances between Populations of *Drosophila Subobscura*, Based on Chromosome Arrangement Frequencies.” *Theoretical and Applied Genetics* 45: 231-41.

Pritchard, JK, M Stephens, and P Donnelly. 2000. “Inference of Population Structure Using Multilocus Genotype Data.” *Genetics* 155: 945-59.

Promislow, DEL, R Montgomerie, and TE Martin. 1992. “Mortality Costs of Sexual Dimorphism in Birds.” *Proceedings of the Royal Society B: Biological Sciences* 250: 143-50.

Purcell, S, B Neale, K Todd-Brown, L Thomas, MAR Ferreira, D Bender, J Maller, P Sklar, PI de Bakker, MJ Daly, et al. 2007. “PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.” *American Journal of Human Genetics* 81: 559-75.

Ratcliffe, N, LR Monteiro, and CJ Hazevoet. 1999. “Status of the Raso Lark *Alauda Razae* with Notes on Threats and Foraging Behaviour.” *Bird Conservation International* 9: 43-6.

Reed, DH. 2005. “Relationship between Population Size and Fitness.” *Conservation Biology* 19: 563-8.

Requedaz, S. 2014. *Découverte Cap-Vert*. Geneva: Editions Olizane.

Reusch, TBH, A Ehlers, A Hammerli, and B Worm. 2005. “Ecosystem Recovery after Climatic Extremes Enhanced by Genotypic Diversity.” *Proceedings of the National Academy of Sciences* 102: 2826-31.

Richardson, DS, FL Jury, DA Dawson, P Salgueiro, J Komdeur, and T Burke. 2000. "Fifty Seychelles Warbler (*Acrocephalus Sechellensis*) Microsatellite Loci Polymorphic in Sylviidae Species and Their Cross-Species Amplification in Other Passerine Birds." *Molecular Ecology* 9: 2226-31.

Ricklefs, RE. 2000. "Intrinsic Aging-Related Mortality in Birds." *Evolution in Health and Disease* 31: 103-11.

Ridgway, R. 1923. "A Plea for Caution in the Use of Trinomials." *The Auk* 40: 375-6.

Ridgway, R. 1919. *The Birds of North and Middle America: A Descriptive Catalogue*. Washington: Government Printing Office.

Rogers, RL, and M Slatkin. 2017. "Excess of Genomic Defects in a Woolly Mammoth on Wrangel Island." *PLoS ONE* 13: e1006601

Romiguier, J, P Gayral, M Ballenghien, A Bernard, V Cahais, A Chenuil, Y Chiari, R Dernat, L Duret, N Faivre, et al. 2014. "Comparative Population Genomics in Animals Uncovers the Determinants of Genetic Diversity." *Nature* 515: 261-3.

Samuel, M, PHF Hobbelen, F Decastro, JA Ahumada, DA Lapointe, CT Atkinson, BL Woodworth, PJ Hart, and DC Duffy. 2011. "The Dynamics, Transmission, and Population Impacts of Avian Malaria in Native Hawaiian Birds: A Modeling Approach." *Ecological Applications* 21: 2960-73.

Santos, ES, and S Nakagawa. 2012. "The Costs of Parental Care: A Meta-Analysis of the Trade-off between Parental Effort and Survival in Birds." *Journal of Evolutionary Biology* 25: 1911-7.

Schluter, D, and J Smith. 1986. "Natural Selection on Beak and Body Size in the Song Sparrow." *Evolution* 40: 221-31.

Schülke, O, and PM Kappeler. 2003. "So Near and yet so Far: Territorial Pairs but Low Cohesion between Pair Partners in a Nocturnal Lemur, *Phaner Furcifer*." *Animal Behaviour* 65: 331-43.

Searcy, WA. 1979. "Sexual Selection and Body Size in Male Red-Winged Blackbirds." *Evolution* 33: 649-61.

Shapiro, LM. 2005. "Diet Overlap and Potential Competition between North Island Brown Kiwi Chicks (*Apteryx Mantelli*) and Ship Rats (*Rattus Rattus*) for Limited Resources on Ponui Island, New Zealand." Massey University. Unpublished PhD thesis.

Sharpe, RB. 1909. *A Hand-List of the Genera and Species of Birds*. London: Taylor and Francis.

Shultz, AJ, AJ Baker, GE Hill, PM Nolan, and SV Edwards. 2016. "SNPs across Time and Space: Population Genomic Signatures of Founder Events and Epizootics in the House Finch (*Haemorhous Mexicanus*)." *Ecology and Evolution* 6: 7475-89.

Simon, JC, TK Pratt, KE Berlin, and JR Kowalsky. 2001. "Reproductive Ecology and Demography of the 'Akohekohe." *Condor* 103: 736-45.

Siriwardena, GM, SR Baillie, and JD Wilson. 1998. "Variation in the Survival Rates of Some British Passerines with Respect to Their Population Trends on Farmland." *Bird Study* 45: 276-92.

Slatkin, M. 2008. "Linkage Disequilibrium--Understanding the Evolutionary Past and Mapping the Medical Future." *Nature Reviews. Genetics* 9: 477-85.

Slavenko, A, OJS Tallowin, Y Itescu, P Raia, S Meiri, and G Thomas. 2016. "Late Quaternary Reptile Extinctions: Size Matters, Insularity Dominates." *Global Ecology and Biogeography* 25: 1308-20.

Smeds, L, A Qvarnström, and H Ellegren. 2016. "Direct Estimate of the Rate of Germline Mutation in a Bird." *Genome Research* 26: 1211-8.

Székely, T, A Liker, RP Freckleton, C Fichtel, and PM Kappeler. 2014. "Sex-Biased Survival Predicts Adult Sex Ratio Variation in Wild Birds." *Proceedings of The Royal Society B: Biological Sciences* 281: 20140342.

Taczanowski, W. 1874. "*Otis Dybowskii*." *Journal for Ornithology* 22: 331.

Tieleman, IB, JB Williams, and P Bloomer. 2003. "Adaptation of Metabolism and Evaporative Water Loss along an Aridity Gradient." *Proceedings of the Royal Society B: Biological Sciences* 270: 207-14.

Tollington, S, CG Jones, A Greenwood, V Tatayah, C Raisin, T Burke, DA Dawson, and JJ Groombridge. 2013. "Long-Term, Fine-Scale Temporal Patterns of Genetic Diversity in the Restored Mauritius Parakeet Reveal Genetic Impacts of Management and Associated Demographic Effects on Reintroduction Programmes." *Biological Conservation* 161: 28-38.

Traill, LW, CJA Bradshaw, and BW Brook. 2007. "Minimum Viable Population Size: A Meta-Analysis of 30 Years of Published Estimates." *Biological Conservation* 139: 159-66.

Tremeau de Rochebrune, A. 1883. *Faune de La Sénégambie I - Atlas. Vertébrés.* Paris: O. Doin.

Wakeley, J. 2008. "Coalescent Theory: An Introduction." *Systematic Biology* 58: 162-5.

Wang, Z, J Zhang, W Yang, N An, P Zhang, G Zhang, and Q Zhou. 2014. “Temporal Genomic Evolution of Bird Sex Chromosomes.” *BMC Evolutionary Biology* 114: 1-12.

Warren, WC, DF Clayton, H Ellegren, AP Arnold, LW Hillier, A Künstner, S Searle, S White, AJ Vilella, S Fairley, et al. 2010. “The Genome of a Songbird.” *Nature* 464: 757-62.

White, GC, and KP Burnham. 1999. “Program MARK: Survival Estimation from Populations of Marked Animals.” *Bird Study* 46: S120-39.

Worley, K, J Collet, LG Spurgin, C Cornwallis, T Pizzari, and DS Richardson. 2010. “MHC Heterozygosity and Survival in Red Junglefowl.” *Molecular Ecology* 19: 3064-75.

Wright, SD, LN Gillman, HA Ross, and DJ Keeling. 2009. “Slower Tempo of Microevolution in Island Birds: Implications for Conservation Biology.” *Evolution* 63: 2275-87.

Wright, AE, R Dean, F Zimmer, and JE Mank. 2016. “How to Make a Sex Chromosome.” *Nature Communications* 7: 12087.

Wyatt, KB, PF Campos, MTP Gilbert, SO Kolokotronis, WH Hynes, R DeSalle, P Daszak, RDE MacPhee, and AD Greenwood. 2008. “Historical Mammal Extinction on Christmas Island (Indian Ocean) Correlates with Introduced Infectious Disease.” *PLoS ONE* 3: e3602.

Supplementary materials

Table 4.S1 Ase18 genotypes of birds sampled between 2004 and 2010 (Brooke et al. 2010; Brooke, Komdeur and van der Velde, personal communication).

Year	Ring number	ASE 18	hex
2004	BN27902	205	224
2006	BN27907	220	224
2005	BN27912	203	220
2006	BN27913	218	224
2007	BN27916	218	224
2004	BN27920	203	224
2006	BN27924	205	220
2004	BN27925	203	224
2006	BN27927	203	205
2006	BN27931	218	218
2008	BN27931	218	218
2006	BN27932	203	224
unknown	BN27933	203	224
2005	BN27935	203	218
2005	BN27936	205	222
2006	BN27940/70625	203	205
2009	BN27945/TJ83316	205	214
2004	BN27947	203	224
2006	BN27951	203	224
2008	BN27951	203	224
2006	BN27953	203	224
2004	BN27954	203	205
2006	BN27955	203	203
2007	BN27957	203	203
2008	BN27957	203	203
2008	BN27961	218	224
2008	BN27966	203	224
2004	BN27967	203	224
2004	BN27968	203	205
2004	BN27969	218	218
2004	BN27970	222	224
2004	BN27971	203	203
2007	BN27971	203	203
2004	BN27972	205	224
2004	BN27973	203	224
2004	BN27974	205	224
2005	BN27976	218	224
2004	BN27977	203	205
2005	BN27983	203	205
2005	BN27984	203	205
2005	BN27985	203	224
2005	BN27986	203	218
2008	BN27986	203	218
unknown	BN27988	203	203
unknown	BN27989	203	224
unknown	BN27990	203	218
2005	BN27991	203	205

2005	BN27992	203	203
unknown	BN27993	220	224
unknown	BN27994	205	224
unknown	BN27995	203	218
2005	BN27996	220	224
2005	BN27997	203	218
2005	BN27998	203	218
2005	BN27999	203	218
unknown	BN28000	203	203
2005	double YL	203	203
2008	double YL	203	203
2005	TC70601	203	203
2005	TC70602	203	203
2005	TC70603	203	224
2005	TC70604	205	218
2005	TC70605	203	205
2005	TC70606	203	224
2005	TC70607	205	220
2005	TC70608	203	224
2005	TC70609	203	224
2005	TC70610	203	218
2005	TC70611	203	224
2005	TC70612	203	203
2005	TC70613	203	203
2005	TC70615	203	205
2005	TC70616	203	205
2005	TC70617	205	218
2005	TC70618	218	224
2006	TC70619	203	203
2006	TC70620	203	203
2006	TC70621	203	205
2006	TC70622	203	205
2006	TC70623	203	203
2006	TC70624	203	224
2006	TC70626	224	224
2006	TC70627	203	224
2006	TC70628	203	205
2006	TC70629	203	205
2006	TC70630	203	205
2006	TC70631	203	224
2006	TC70632	203	224
2006	TC70633	203	205
2006	TC70634	224	224
2006	TC70635	203	205
2006	TC70636	203	220
2006	TC70637	203	224
2006	TC70638	203	203
2006	TC70639	214	224
2006	TC70640	205	218
2006	TC70641	203	218
2006	TC70642	205	224
2006	TC70643	203	203
2006	TC70644	205	220
2006	TC70645	203	224
2006	TC70646	205	224
2006	TC70647	203	203
2007	TC70650	203	205
2007	TC70651	224	224
2007	TC70652	203	224
2007	TC70653	205	220

2007	TC70654	205	218
2007	TC70655	203	224
2007	TC70656	203	218
2007	TC70657	205	218
2007	TC70658	205	224
2007	TC70659	203	218
2007	TC70660	203	205
2007	TC70661	205	224
2007	TC70662	203	205
2007	TC70663	203	205
2007	TC70664	203	224
2007	TC70665	203	205
2007	TC70666	203	205
2007	TC70667	205	224
2007	TC70668	205	224
2007	TC70669	205	224
2008	TC70670	203	205
2008	TC70671	218	218
2008	TC70672	203	220
2008	TC70673	218	224
2008	TC70674	218	224
2008	TC70675	203	205
2008	TC70676	203	205
2008	TC70677	218	224
2008	TC70678	203	218
2010	TC70678	203	218
2008	TC70679	203	205
2008	TC70680	203	205
2008	TC70681	205	224
2008	TC70683	203	203
2008	TC70684	203	203
2008	TC70685	unknown	unknown
2008	TC70686	203	224
2008	TC70687	203	203
2008	TC70688	203	205
2008	TC70689	203	203
2008	TC70690	203	205
2008	TC70692	unknown	unknown
2009	TC70693	205	224
2009	TC70694	205	218
2009	TC70695	203	203
2009	TC70696	205	224
2009	TC70697	203	205
2009	TC70698	203	224
2009	TC70699	203	224
2009	TC70700	203	224
2009	TJ83301	203	205
2009	TJ83302	218	224
2009	TJ83303	218	224
2009	TJ83304	205	224
2009	TJ83305	203	224
2009	TJ83306	205	224
2009	TJ83307	218	224
2009	TJ83308	205	218
2009	TJ83309	205	218
2009	TJ83310	218	224
2009	TJ83311	203	205
2009	TJ83312	203	220
2009	TJ83313	218	224
2009	TJ83314	205	224

2009	TJ83315	203	224
2009	TJ83317	203	203
2009	TJ83318	203	218
2009	TJ83319	unknown	unknown
2009	TJ83320	203	218
2009	TJ83321	203	205
2009	TJ83322	218	224
2009	TJ83323	203	224
2009	TJ83324	203	224
2009	TJ83325	203	205
2009	TJ83326	203	205
2009	TJ83328	203	205
2009	TJ83329	203	205
2009	TJ83330	203	205
2009	TJ83331	203	224
2009	TJ83332	203	224
2009	TJ83333	203	218
2009	TJ83334	203	224
2009	TJ83335	205	224
2009	TJ83336	203	224
2009	TJ83337	205	218
2009	TJ83339	203	205
2009	TJ83340	203	205
2009	TJ83341	203	205
2009	TJ83343	203	205
2010	TJ83344	203	224
2010	TJ83345	203	205
2010	TJ83346	203	205
2010	TJ83347	203	224
2010	TJ83348	218	224
2010	TJ83349	205	224
2010	TJ83350	203	224
2010	TJ83351	205	224
2010	TJ83352	203	205
2010	TJ83353	203	224
2010	TJ83354	203	203
2010	TJ83355	203	224
2010	TJ83356	203	205
2010	TJ83357	205	224
2010	TJ83358	203	224
2010	TJ83359	203	203
2010	TJ83360	203	205
2010	TJ83361	205	218
2010	TJ83362	203	224
2010	TJ83363	203	224
2010	TJ83364	218	218
2010	TJ83365	203	203
2010	TJ83366	203	205
2010	TJ83367	205	220
2010	TJ83368	203	205
2010	TJ83369	224	224
2010	TJ83370	203	205
2010	TJ83371	205	224
2010	TJ83372	203	218
2010	TJ83373	203	205
2010	TJ83374	218	218
2010	TJ83375	205	224
2010	TJ83376	218	224
2010	TJ83377	224	224
2010	TJ83378	203	205

2010	TJ83379	205	224
2010	TJ83380	203	205
2010	TJ83381	205	218
2010	TJ83382	205	218
2010	TJ83383	203	218
2010	TJ83384	203	205
2010	TJ83385	205	224
2010	TJ83386	203	224
2010	TJ83387	203	205
2010	TJ83388	224	224
2010	TJ83389	203	205
2010	TJ83390	203	218
2010	TJ83391	203	203
2010	TJ83392	205	224
2010	TJ83393	203	203
2010	TJ83394	203	205
2010	TJ83395	203	203
2010	TJ83396	224	224
2010	TJ83397	203	218
2010	TJ83398	203	205
2010	TJ83399	203	218
2010	TJ83400	203	205
2010	TS05201	205	220
2010	TS05202	205	224
2010	TS05203	205	218
2010	TS05204	218	224
2010	TS05205	203	205
2010	TS05206	205	224
2010	TS05207	203	203
2010	TS05208	203	203
2010	TS05209	203	224
2010	TS05210	203	205
2010	TS05211	203	205
2010	TS05212	203	205
2010	TS05213	205	224
2010	TS05214	203	203
2010	TS05215	205	224
2010	TS05216	203	224
2010	TS05217	203	205
2010	TS05218	203	205
2010	TS05219	203	224
2010	TS05220	205	224
2010	TS05221	203	224
2010	TS05222	224	224
2010	TS05223	205	224
2010	TS05224	203	224
2010	TS05225	205	224
2010	TS05226	203	205
2010	TS05227	203	205
2010	TS05228	205	224
2010	TS05229	205	224
2010	TS05230	203	218
2010	TS05231	218	224
2010	TS05232	203	224
2010	TS05233	218	224
2010	TS05234	205	218
2010	TS05235	203	205
2010	TS05236	205	218
2010	TS05237	205	218
2010	TS05238	203	224

2010	TS05239	205	224
2010	TS05240	224	224
2010	TS05241	205	224
2010	TS05242	203	205
2010	TS05243	205	224
2010	TS05246	203	224
2010	TS05247	205	218
2010	TS05248	205	224
2010	TS05249	203	205
2010	TS05250	205	224
2010	TS05251	205	224
2010	TS05252	203	224
2010	TS05253	205	224
2010	TS05254	205	218
2010	TS05255	203	203
2010	TS05256	203	203
2010	TS05257	203	205
2010	TS05258	203	205
2010	TS05259	205	224
2010	TS05260	203	203
2010	TS05261	214	224
2010	TS05262	205	224
2010	TS05263	214	224
2010	TS05264	205	220
2010	TS05265	218	224

Table 5.S1 Samples used for the genetic analyses in Chapters 5 and 6.

Sample number (analyses)	Sample number (on tube)	Species	Sex	Collection date	Country	Site	Type	Provenance
0	Harvard MCZ 349891	*Alauda arvensis*	male	14/06/12	Mongolia	Hustai National Park	tissue	Harvard MCZ
1	70645	*Alauda razae*	male	15/11/06	Cape Verde	Raso	blood	M. Brooke
2	13	*Alauda arvensis*	unknown	11/06/14	Netherlands	Aekingerzand	red blood cells	P. van Veelen
3	2	*Alauda arvensis*	unknown	17/05/13	Netherlands	Aekingerzand	red blood cells	P. van Veelen
4	5	*Alauda arvensis*	unknown	29/05/14	Netherlands	Aekingerzand	red blood cells	P. van Veelen
5	TT95628	*Alauda razae*	male	14/11/13	Cape Verde	Raso	blood	M. Brooke
6	11	*Alauda arvensis*	unknown	18/06/14	Netherlands	Aekingerzand	red blood cells	P. van Veelen
7	TT95651	*Alauda razae*	female	18/11/13	Cape Verde	Raso	blood	M. Brooke
8	3	*Alauda arvensis*	unknown	23/05/14	Netherlands	Aekingerzand	red blood cells	P. van Veelen
9	7	*Alauda arvensis*	unknown	20/06/14	Netherlands	Aekingerzand	red blood cells	P. van Veelen
10	4	*Alauda arvensis*	unknown	21/05/14	Netherlands	Aekingerzand	red blood cells	P. van Veelen
11	12	*Alauda arvensis*	unknown	13/06/14	Netherlands	Aekingerzand	red blood cells	P. van Veelen
12	6	*Alauda arvensis*	unknown	15/05/14	Netherlands	Aekingerzand	red blood cells	P. van Veelen
13	70650	*Alauda razae*	female	09/11/07	Cape Verde	Raso	blood	M. Brooke
14	8	*Alauda arvensis*	unknown	30/04/2014	Netherlands	Aekingerzand	red blood cells	P. van Veelen
15	TT95686	*Alauda razae*	male	26/11/14	Cape Verde	Raso	blood	M. Brooke
16	TT95652	*Alauda razae*	male	18/11/13	Cape Verde	Raso	blood	M. Brooke
17	TT95621	*Alauda razae*	female	12/11/13	Cape Verde	Raso	blood	M. Brooke
18	1	*Alauda arvensis*	unknown	16/05/13	Netherlands	Aekingerzand	red blood cells	P. van Veelen
19	15	*Alauda arvensis*	unknown	20/06/14	Netherlands	Aekingerzand	red blood cells	P. van Veelen
20	TT95663	*Alauda razae*	female	24/11/13	Cape Verde	Raso	blood	M. Brooke
21	TT95658	*Alauda razae*	male	20/11/13	Cape Verde	Raso	blood	M. Brooke
22	TT95607	*Alauda razae*	male	09/11/13	Cape Verde	Raso	blood	M. Brooke
23	TT95609	*Alauda razae*	male	09/11/13	Cape Verde	Raso	blood	M. Brooke
24	TT95650	*Alauda razae*	male	18/11/13	Cape Verde	Raso	blood	M. Brooke
25	TT95684	*Alauda razae*	male	20/11/14	Cape Verde	Raso	blood	M. Brooke
26	TT95660	*Alauda razae*	female	21/11/13	Cape Verde	Raso	blood	M. Brooke
27	TT95600	*Alauda razae*	male	08/11/13	Cape Verde	Raso	blood	M. Brooke
28	10	*Alauda arvensis*	unknown	18/06/14	Netherlands	Aekingerzand	red blood cells	P. van Veelen
29	TT95625	*Alauda razae*	female	13/11/13	Cape Verde	Raso	blood	M. Brooke
30	TT95691	*Alauda razae*	female	20/11/14	Cape Verde	Raso	blood	M. Brooke
31	14	*Alauda arvensis*	unknown	11/06/14	Netherlands	Aekingerzand	red blood cells	P. van Veelen

32	9	*Alauda arvensis*	unknown	16/09/14	Netherlands	Aekingerzand	red blood cells	P. van Veelen
33	SYSb1479	*Alauda gulgula*	male	06/12/05	Taiwan	Taichung	muscle	Y. Liu
34	SYSb1490	*Alauda gulgula*	unknown	29/01/02	Taiwan	Hualian	muscle	Y. Liu
35	SYSb1486	*Alauda gulgula*	male	06/08	Taiwan	Taipei	muscle	Y. Liu
36	SYSb1485	*Alauda gulgula*	male	03/05/07	Taiwan	Taichung	muscle	Y. Liu
37	SYSb1488	*Alauda gulgula*	male	14/05/02	Taiwan	Taichung	muscle	Y. Liu
38	83322	*Alauda razae*	male	20/11/09	Cape Verde	Raso	blood	M. Brooke
39	SYSb1480	*Alauda gulgula*	unknown	23/10/07	Taiwan	Taichung	muscle	Y. Liu
40	TT95674	*Alauda razae*	female	15/11/14	Cape Verde	Raso	blood	M. Brooke
41	SYSb000799	*Alauda arvensis*	female	21/06/94	Russia	Sverdlovskaya Oblast	muscle	Y. Liu
42	CL2	*Galerida cristata*	female	27/05/07	Saudi Arabia	Taif	red blood cells	N. Horrocks
43	SYSb1477	*Alauda gulgula*	male	24/05/02	Taiwan	Taichung	muscle	Y. Liu
44	SYSb000790	*Alauda arvensis*	male	23/05/94	Russia	Moscovskaya Oblast	muscle	Y. Liu
45	SYSb000792	*Alauda arvensis*	female	22/05/94	Russia	Moscovskaya Oblast	muscle	Y. Liu
46	TT95680	*Alauda razae*	female	19/11/14	Cape Verde	Raso	blood	M. Brooke
47	CL1	*Galerida cristata*	female	27/05/07	Saudi Arabia	Taif	red blood cells	N. Horrocks
48	CL7	*Galerida cristata*	male	06/12/07	Saudi Arabia	Mahazat as-Sayd	red blood cells	N. Horrocks
49	TT95683	*Alauda razae*	male	20/11/14	Cape Verde	Raso	muscle	M. Brooke
50	SYSb000791	*Alauda arvensis*	female	22/05/94	Russia	Moscovskaya Oblast	muscle	Y. Liu
51	CL4	*Galerida cristata*	female	28/05/07	Saudi Arabia	Taif	red blood cells	N. Horrocks
52	CL3	*Galerida cristata*	female	27/05/07	Saudi Arabia	Taif	red blood cells	N. Horrocks
53	SYSb1478	*Alauda gulgula*	unknown	28/06/01	Taiwan	Hualian	muscle	Y. Liu
54	SYSb1481	*Alauda gulgula*	male	26/11/07	Taiwan	Taichung	muscle	Y. Liu
55	SYSb1476	*Alauda gulgula*	male	26/11/07	Taiwan	Taichung	muscle	Y. Liu
56	CL5	*Galerida cristata*	female	02/12/07	Saudi Arabia	Mahazat as-Sayd	red blood cells	N. Horrocks
57	SYSb000795	*Alauda arvensis*	male	21/06/94	Russia	Sverdlovskaya Oblast	muscle	Y. Liu
58	SYSb1482	*Alauda gulgula*	male	06/12/05	Taiwan	Taichung	muscle	Y. Liu
59	SYSb000793	*Alauda arvensis*	male	23/05/94	Russia	Moscovskaya Oblast	muscle	Y. Liu
60	SYSb000798	*Alauda arvensis*	female	21/06/94	Russia	Sverdlovskaya Oblast	muscle	Y. Liu
61	SYSb000797	*Alauda arvensis*	male	21/06/94	Russia	Sverdlovskaya Oblast	muscle	Y. Liu
62	CL8	*Galerida cristata*	female	11/12/07	Saudi Arabia	Mahazat as-Sayd	red blood cells	N. Horrocks
63	SYSb000800	*Alauda arvensis*	male	21/06/94	Russia	Sverdlovskaya Oblast	muscle	Y. Liu
64	SYSb000794	*Alauda arvensis*	male	04/06/94	Russia	Kurskaya Oblast	muscle	Y. Liu
65	SVD2472	*Alauda arvensis*	female	28/07/01	Russia	Primorskiy Kray, Spasskiy Rayon, Gayvoron	muscle	Y. Liu
66	AWH184	*Alauda arvensis*	male	15/07/02	Russia	Ulan-Ude	muscle	Y. Liu
67	unknown	*Alauda razae*	unknown	unknown	Cape Verde	Raso	blood	M. Brooke
68	SYSb059	*Alauda arvensis*	female	17/07/08	China	Qinghai	muscle	Y. Liu
69	SYSb2937	*Alauda arvensis*	male	?/07/11	China	Qinghai	blood	Y. Liu
70	70666	*Alauda razae*	female	16/11/07	Cape Verde	Raso	blood	M. Brooke
71	70646	*Alauda razae*	female	18/11/06	Cape Verde	Raso	blood	M. Brooke

72	70643	*Alauda razae*	male	14/11/06	Cape Verde	Raso	blood	M. Brooke
73	70664	*Alauda razae*	male	14/11/07	Cape Verde	Raso	blood	M. Brooke
74	646	*Alauda arvensis*	female	17/07/08	China	Qinghai	DNA	Y. Liu
75	JMB1217	*Alauda arvensis*	male	03/08/92	Russia	Magadanskaya Oblast	muscle	Y. Liu
76	double yellow	*Alauda razae*	male	08/12/08	Cape Verde	Raso	blood	M. Brooke
77	70610	*Alauda razae*	male	12/12/05	Cape Verde	Raso	blood	M. Brooke
78	27957	*Alauda razae*	male	06/12/08	Cape Verde	Raso	blood	M. Brooke
79	FwXYQ	*Alauda arvensis*	unknown	summer	China	Qinghai	muscle	Y. Liu
80	70659	*Alauda razae*	male	13/11/07	Cape Verde	Raso	blood	M. Brooke
81	70657	*Alauda razae*	female	12/11/07	Cape Verde	Raso	blood	M. Brooke
82	JMB1163	*Alauda arvensis*	female	25/07/92	Russia	Kamchatka, Oktyabr'skiy	muscle	Y. Liu
83	589	*Alauda arvensis*	unknown	20/08/02	China	Qinghai	DNA	Y. Liu
84	SVD527	*Alauda arvensis*	female	10/06/94	Russia	Irkutskaya Oblast	muscle	Y. Liu
85	SAR6386	*Alauda arvensis*	male	15/06/93	Russia	Nizhnetambovskoye	muscle	Y. Liu
86	China1	*Alauda arvensis*	unknown	summer	China	Qinghai	muscle	Y. Liu
87	China2	*Alauda arvensis*	unknown	summer	China	Qinghai	muscle	Y. Liu
88	CSW5819	*Alauda arvensis*	female	17/05/98	Mongolia	Choibalsan	muscle	Y. Liu
89	70663	*Alauda razae*	female	14/11/07	Cape Verde	Raso	blood	M. Brooke
90	BKS3905	*Alauda arvensis*	female	22/05/97	Mongolia	Töv Aymag	muscle	Y. Liu
91	CL6	*Galerida cristata*	male	04/12/07	Saudi Arabia	Mahazat as-Sayd	red blood cells	N. Horrocks
92	KIL74	*Alauda arvensis*	male	15/06/03	Russia	Sakhalinskaya Oblast	muscle	Y. Liu
93	CL9	*Galerida cristata*	male	12/12/07	Saudi Arabia	Mahazat as-Sayd	red blood cells	N. Horrocks
94	83364	*Alauda razae*	male	11/11/10	Cape Verde	Raso	blood	M. Brooke
95	China3	*Alauda arvensis*	unknown	summer	China	Qinghai	muscle	Y. Liu
96	BKS1065	*Alauda arvensis*	female	15/07/93	Russia	Sakhalinskaya Oblast	muscle	Y. Liu

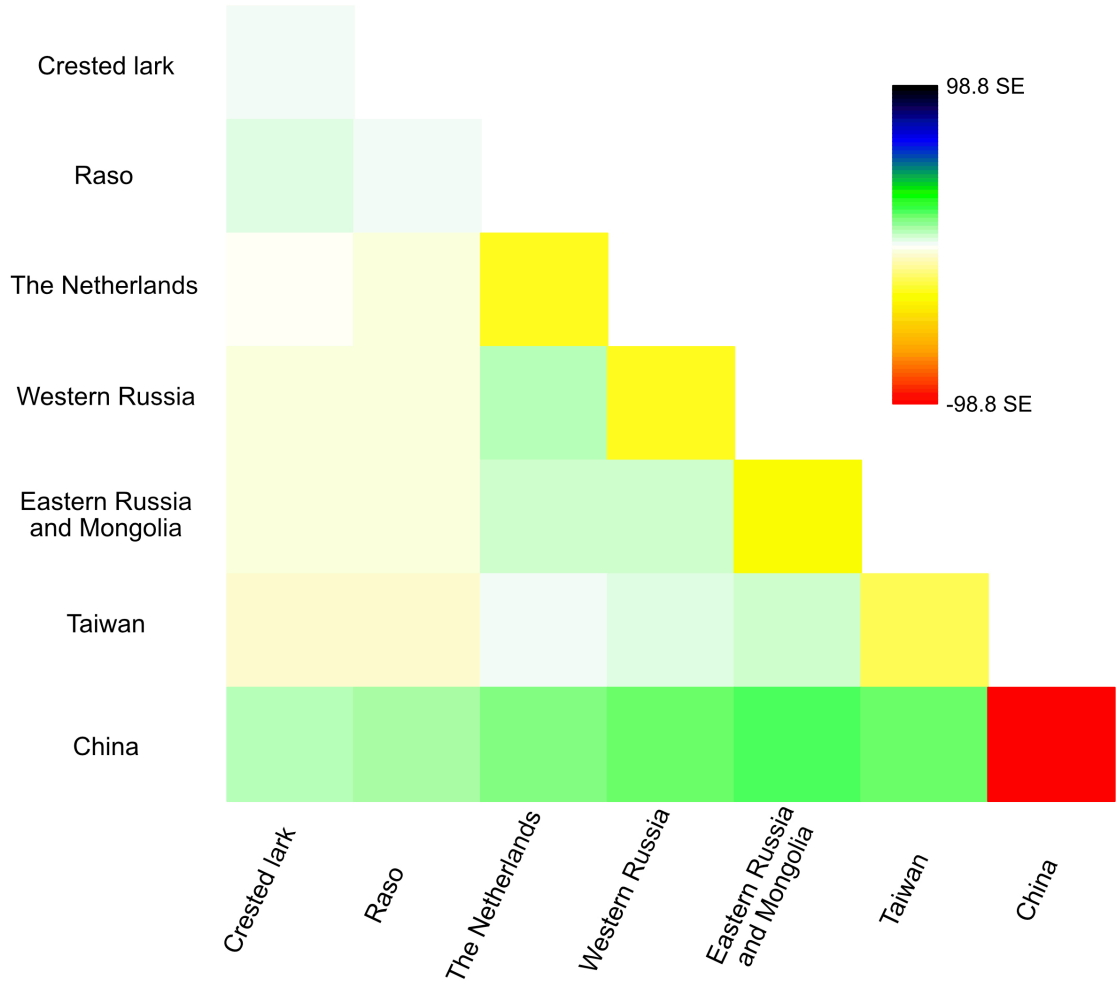


Figure 5.S1 Residuals for Figure 5.7, as outputted by TreeMix. The residual covariance between each pair of populations is divided by the average standard error across all pairs. This scaled residual is then plotted in each cell.